



International
Institute of Information
Technology Bangalore

Lending Club Case Study

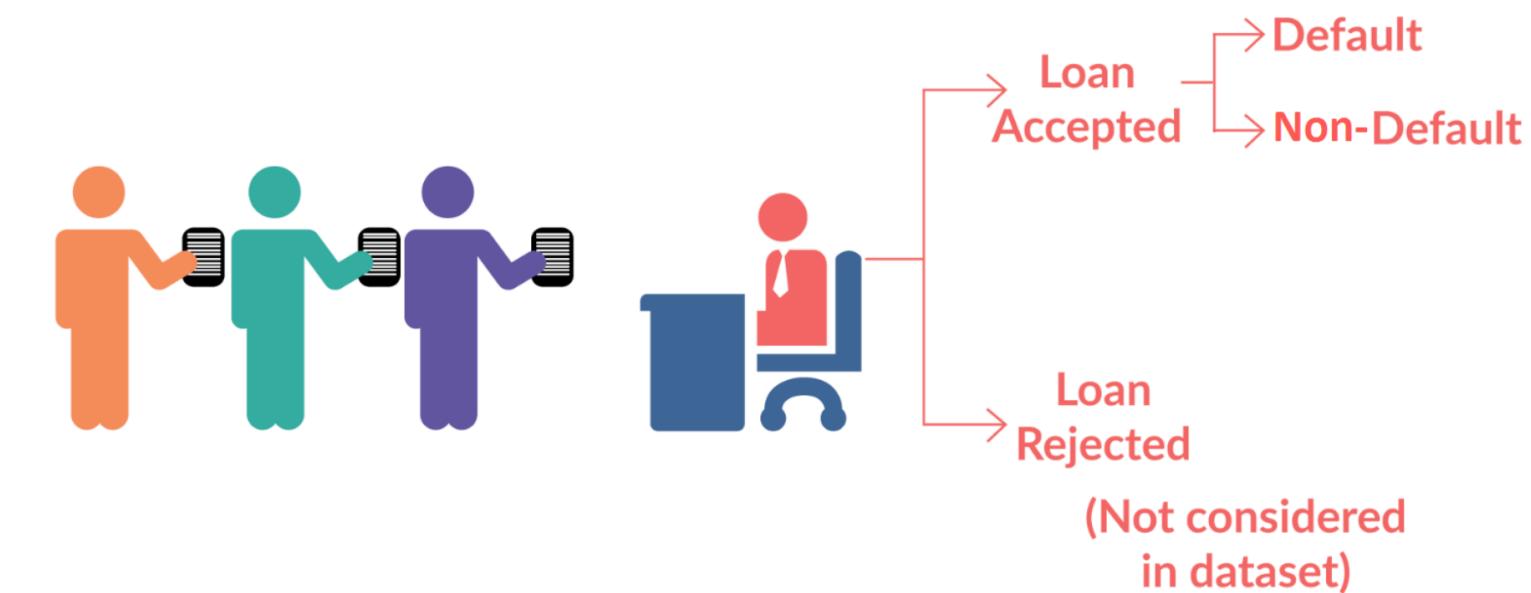
Exploratory Data Analysis

Group Facilitator : Shashank Bhatnagar
Group Member : Aayushi Mittal

Table of Contents

- Problem Statement
- Business Objective and Aim
- Exploratory Data Analysis Methodology
- Analysis
- Conclusion

Problem Statement



Working for a consumer finance company which specialises in lending various types of loans to urban customers. When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile. 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

When a person applies for a loan, there are two types of decisions that could be taken by the company:

Loan accepted: If the company approves the loan, there are 3 possible scenarios described below:

- Fully paid: Applicant has fully paid the loan (the principal and the interest rate)
- 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

Loan rejected: The company had rejected the loan (because the candidate does not meet their requirements etc.). Since the loan was rejected, there is no transactional history of those applicants with the company and so this data is not available with the company (and thus in this dataset)

Business Objective & Aim

Like most other lending companies, lending loans to 'risky' applicants is the largest source of financial loss (called credit loss). The credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed. In other words, borrowers who default cause the largest amount of loss to the lenders. In this case, the customers labelled as 'charged-off' are the 'defaulters'.

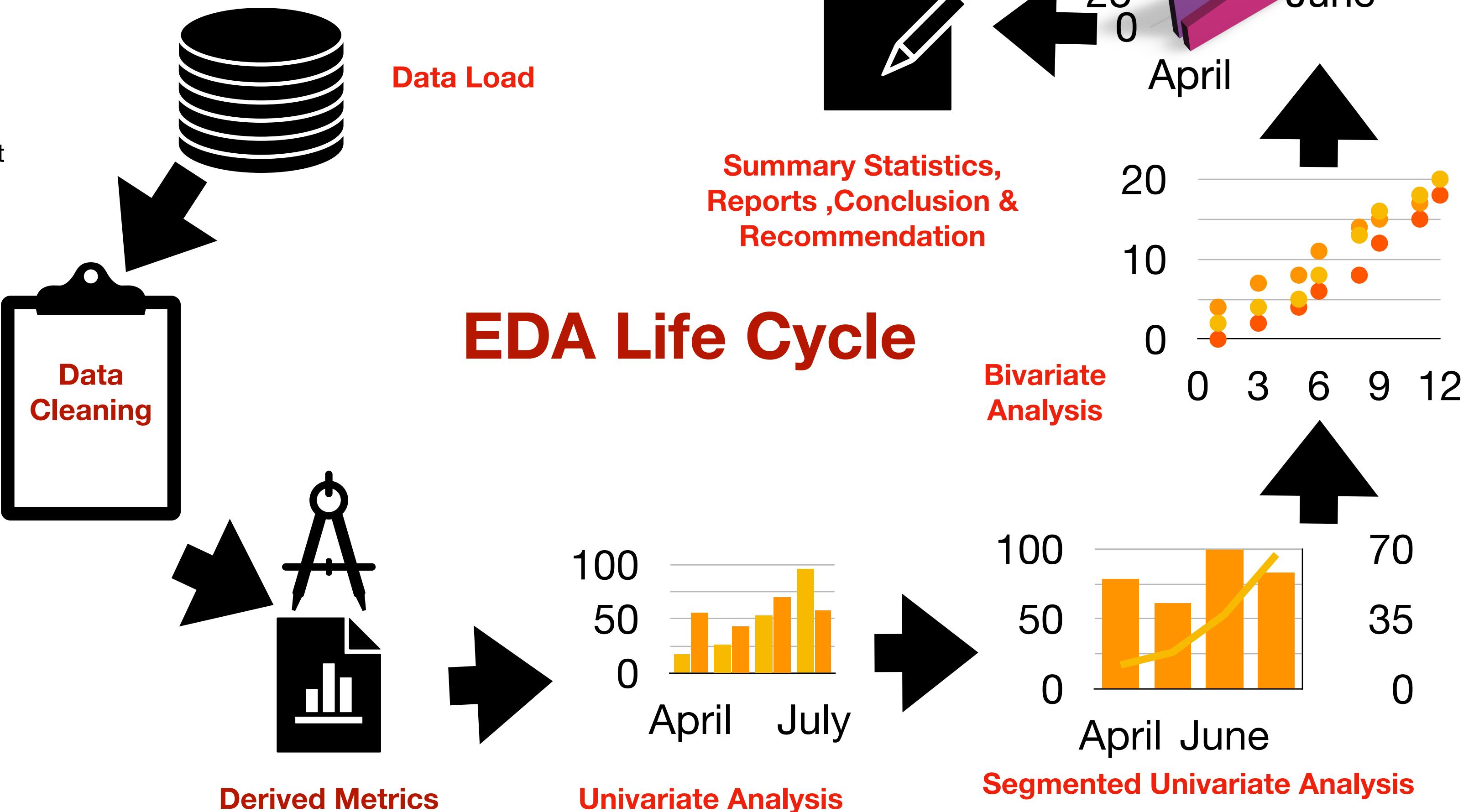
- The company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default.
- The company can utilise this knowledge for its portfolio and risk assessment.
- If one is able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. Identification of such applicants using EDA is the aim of this case study.
- Our aim is to identify the loan application which can be defaulted or delinquent 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

Dataset : Loan data 2007 - 2011

Exploratory Data Analysis and Methodology

Steps

- Data Cleaning Missing Data Treatment
- Data Cleaning Removing Customer Behavior Variables
- Data Cleaning - Standardizing Values, Units, Date & Text
- Data Cleaning - Outlier Analysis
- Derived Metrics & Binning
- Univariate Analysis
 - Continuous variable
 - Categorical variable
- Segmented Univariate Analysis
- Bivariate Analysis
 - Continuous vs Continuous
 - Continuous vs Categorical
 - Categorical vs Categorical
- MultiVariate Analysis & Correlation Matrix
- Summary Statistics
- Missing Data Treatment Analysis Report
- Univariate Data Analysis Report
- Bivariate Data Analysis Report
- Final Conclusion



Analysis

Highlights:

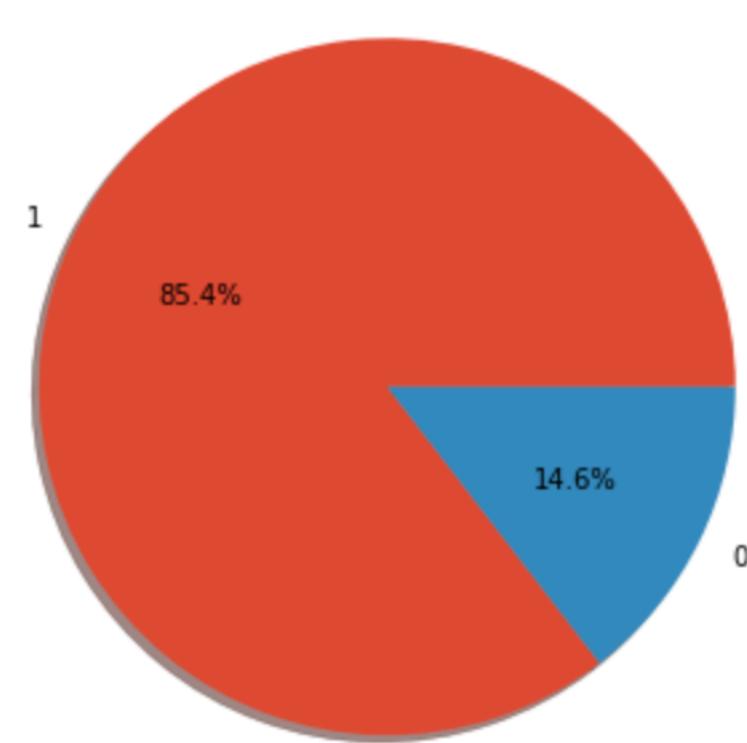
15 percent of total population are defaulted

The Most of loan applied by the Customers are usually living on Rent i.e 47.9% and then having Mortgage

Around 44 % of loan application are in Not Verified Status

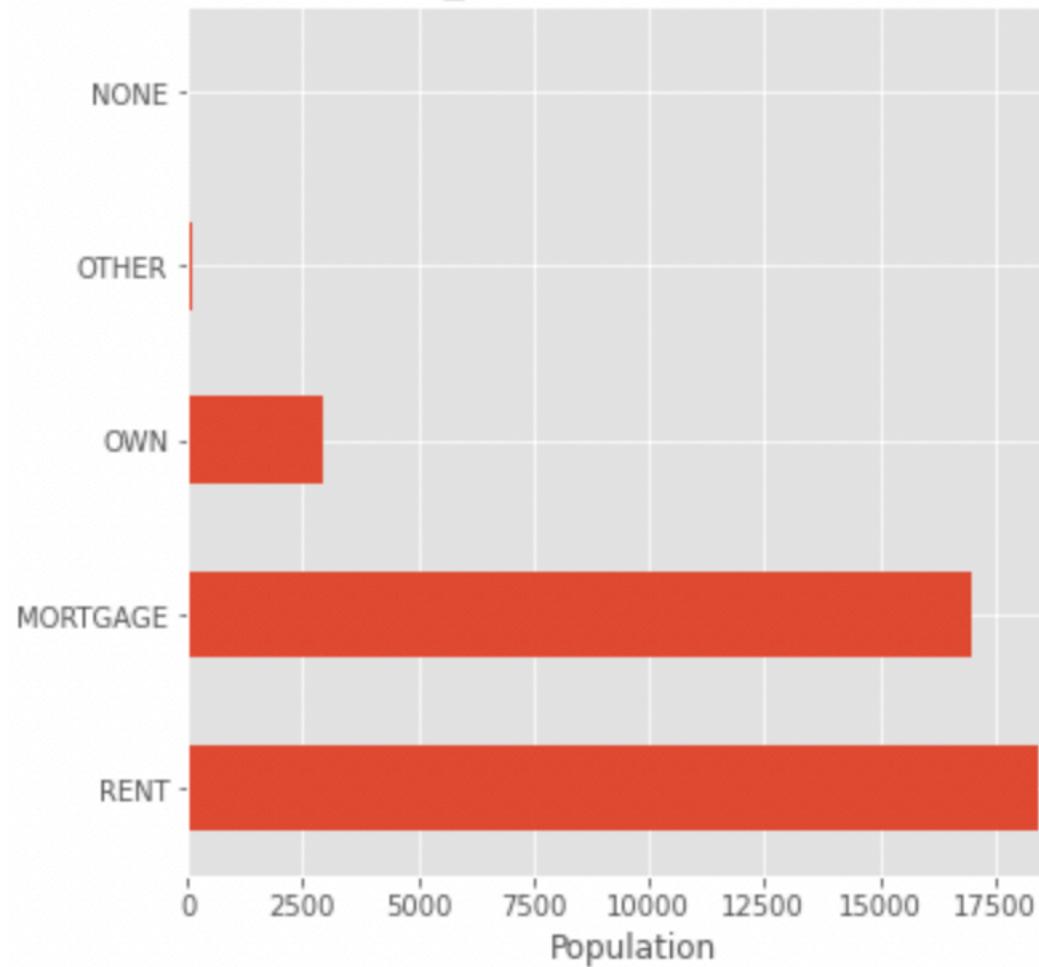
Around 60% of loans are take for Debt Consolidation & for Credit Card Purpose

loan_status vs population %

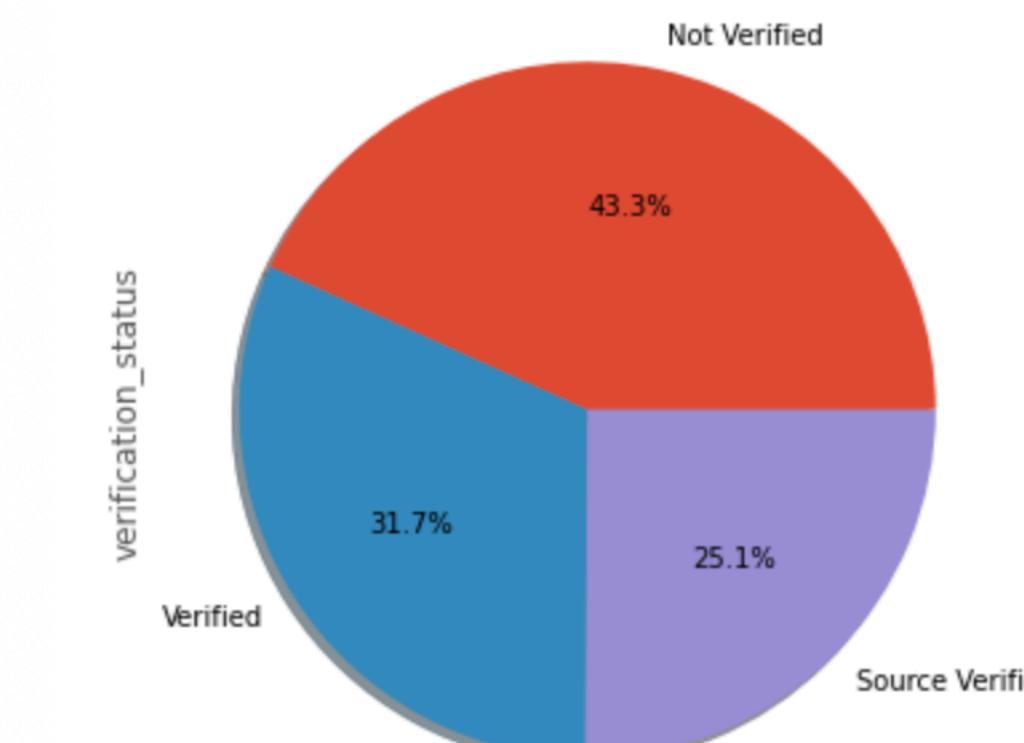


0 = "Defaulted"
1 = "Fully Paid"

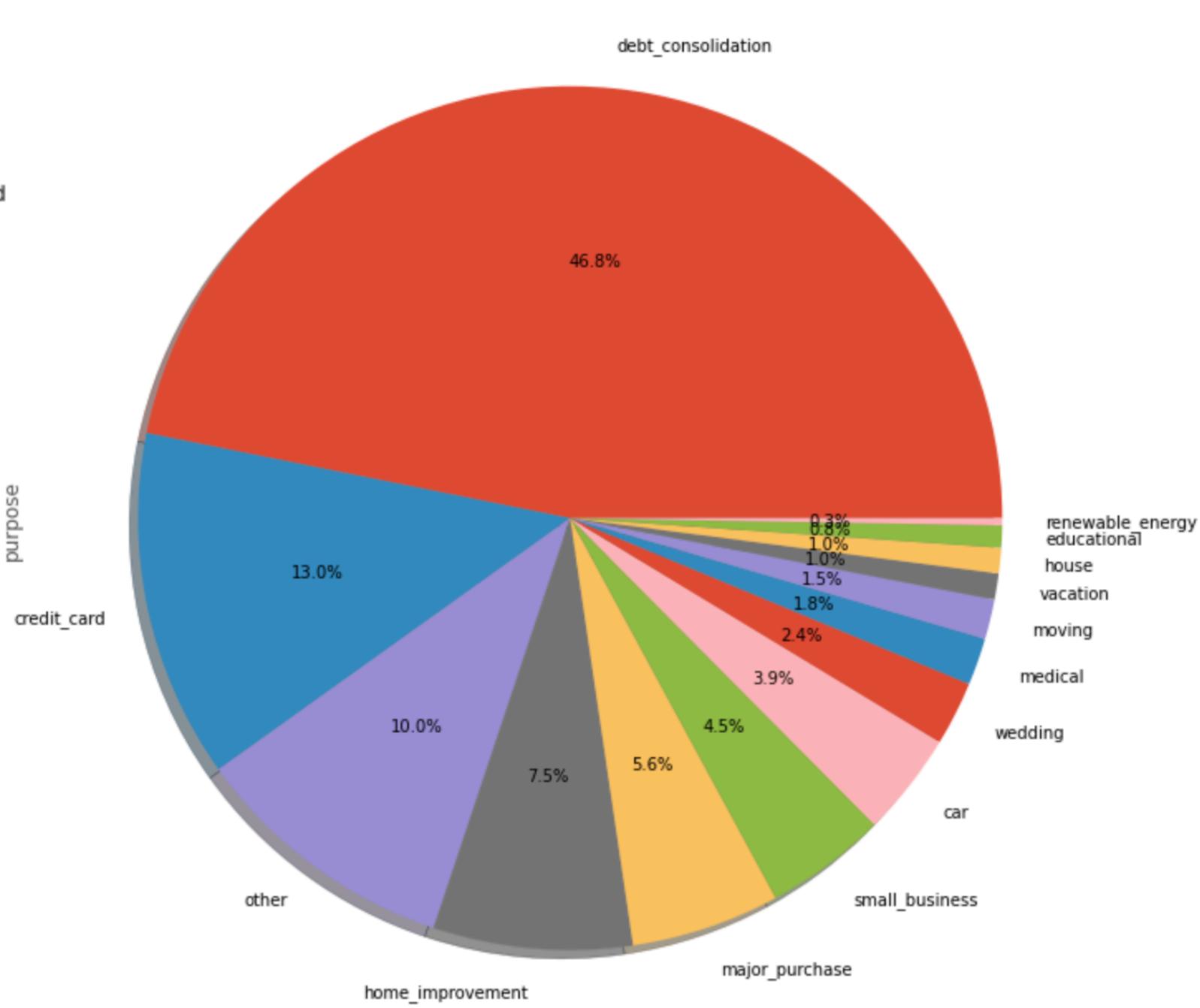
home_ownership vs population



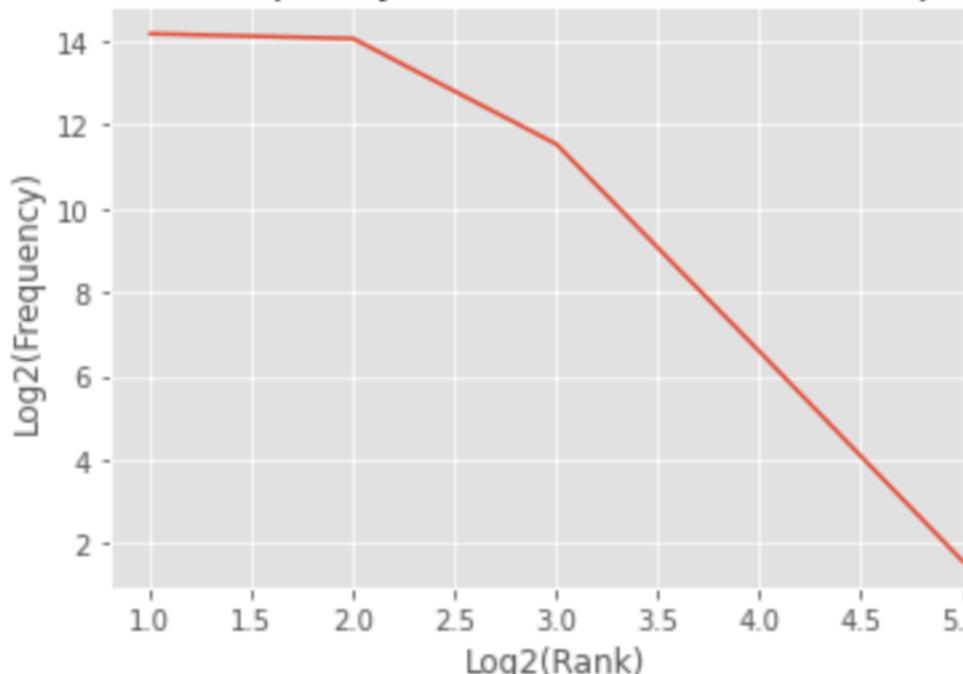
verification_status vs population %



purpose vs population %



Rank vs Frequency - Power Law Distribution - purpose



Homeownership PowerLaw distribution

Highlights:



International
Institute of Information
Technology Bangalore

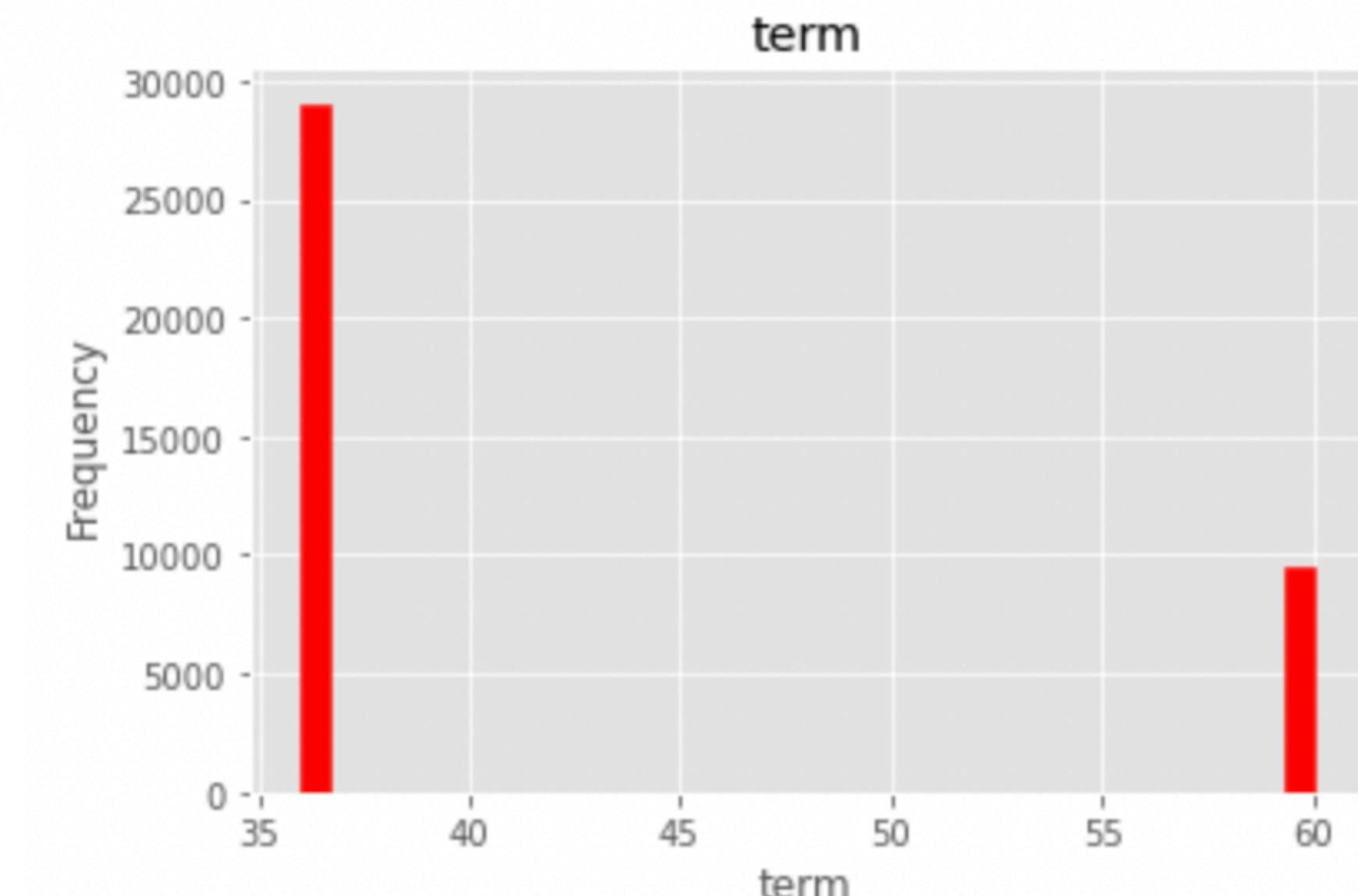
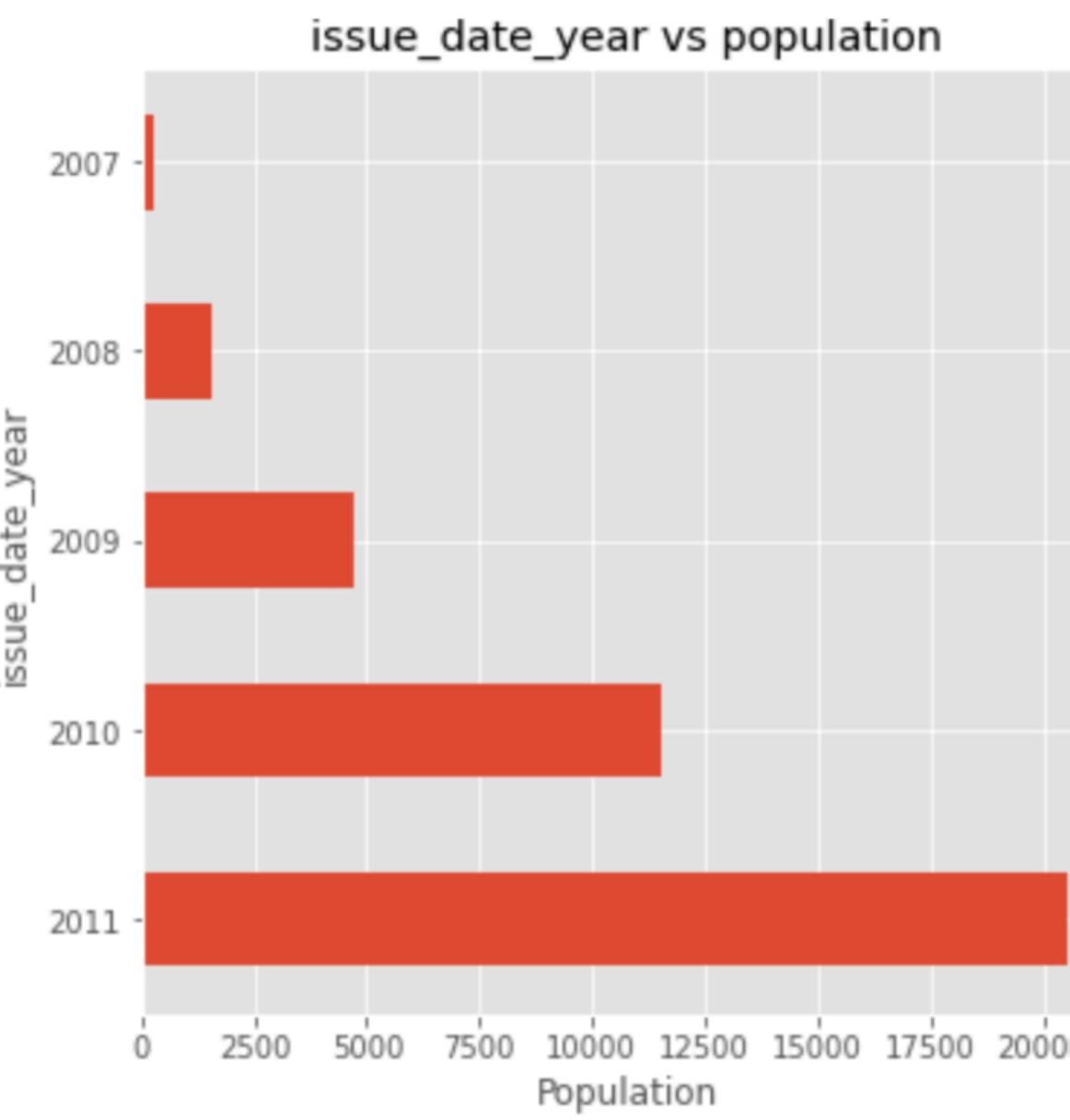
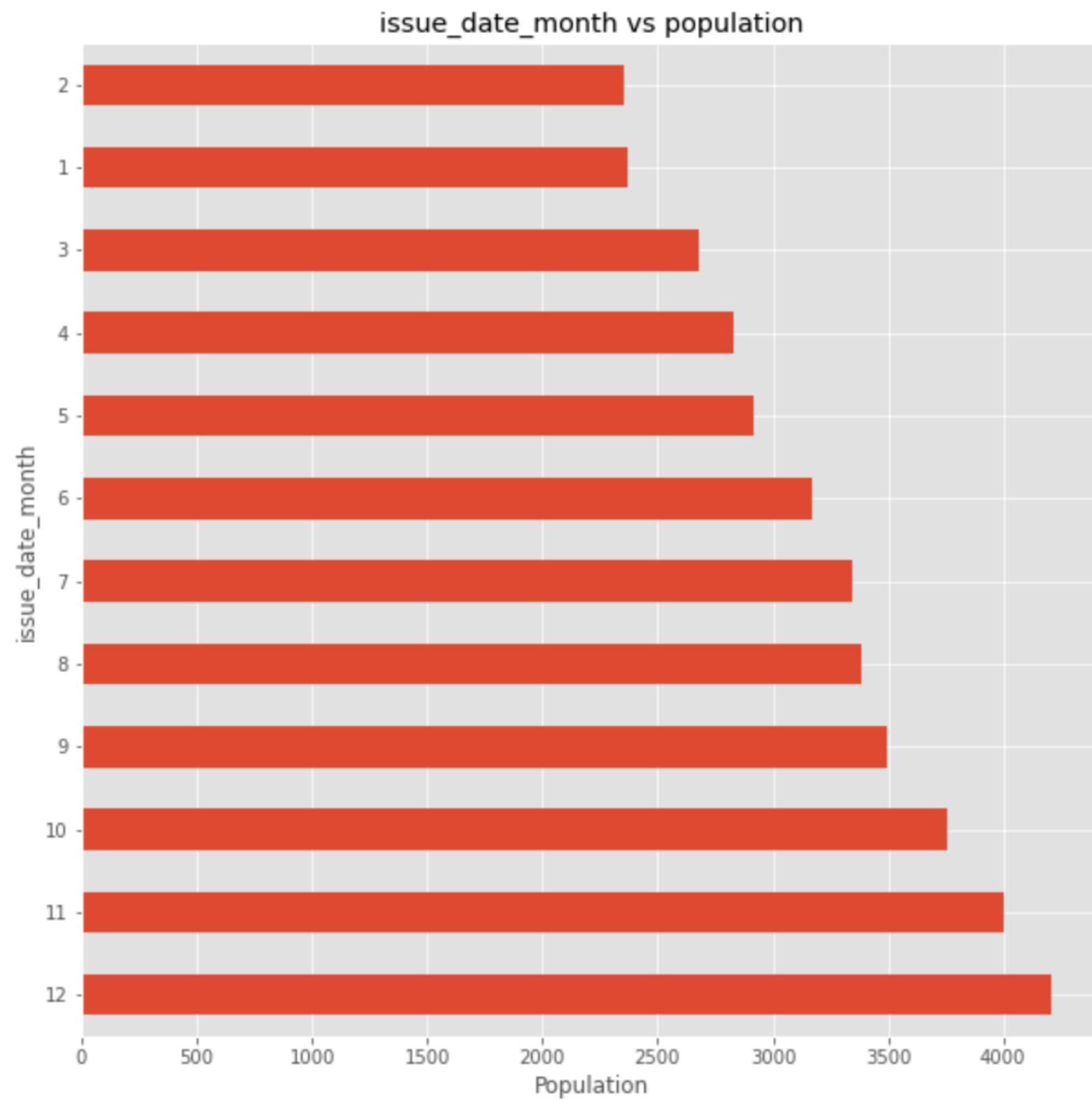
Analysis

Lending Club has been increased year by year from 2007 to 2011

Can see a continuous decline in number of loan application from Dec to Jan

Maximum no of loan applications are taken for term 36 months

Around 25 % of loan application have Debt to income ratio greater than 25 % - These customers can be risky



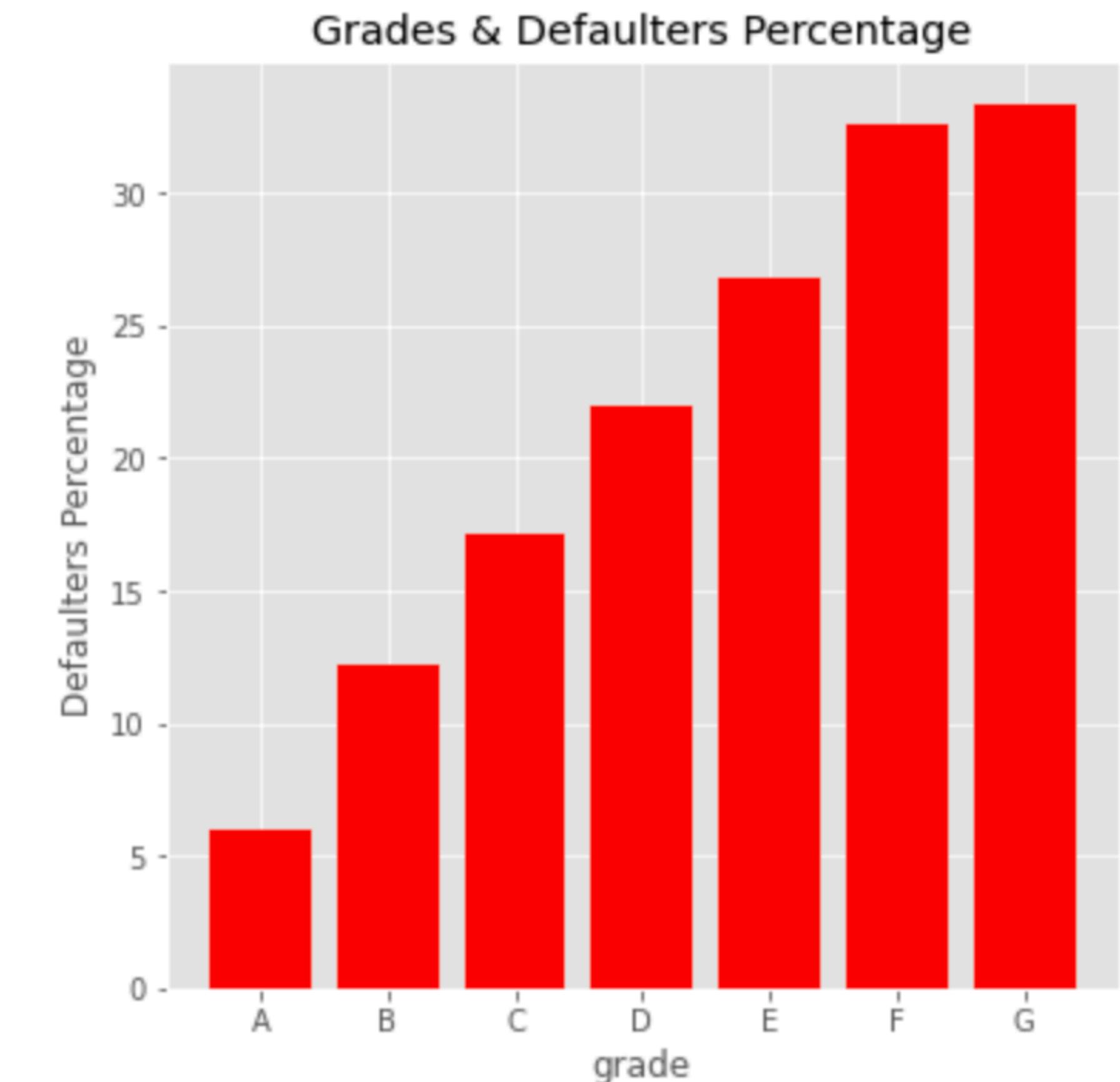
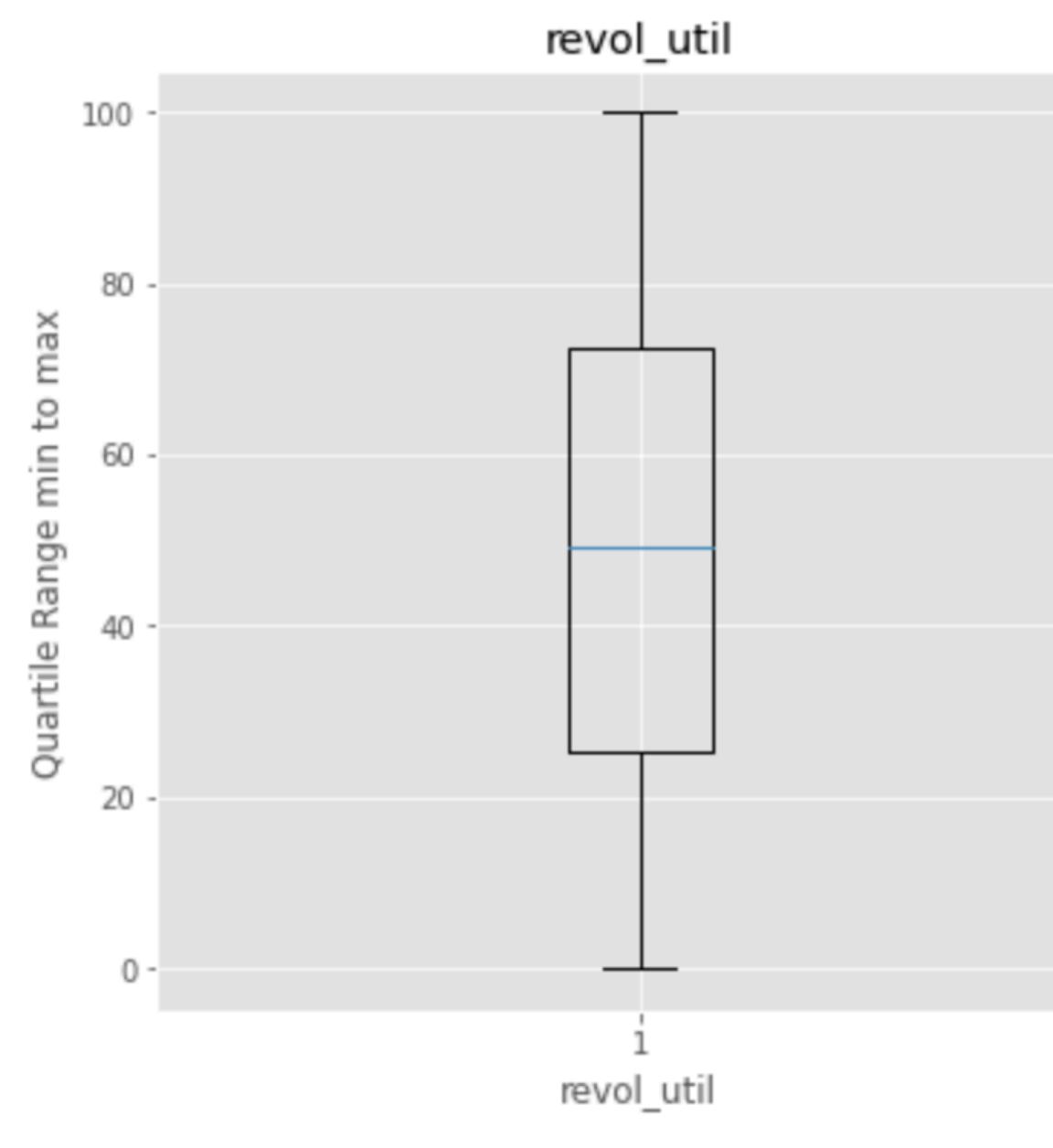
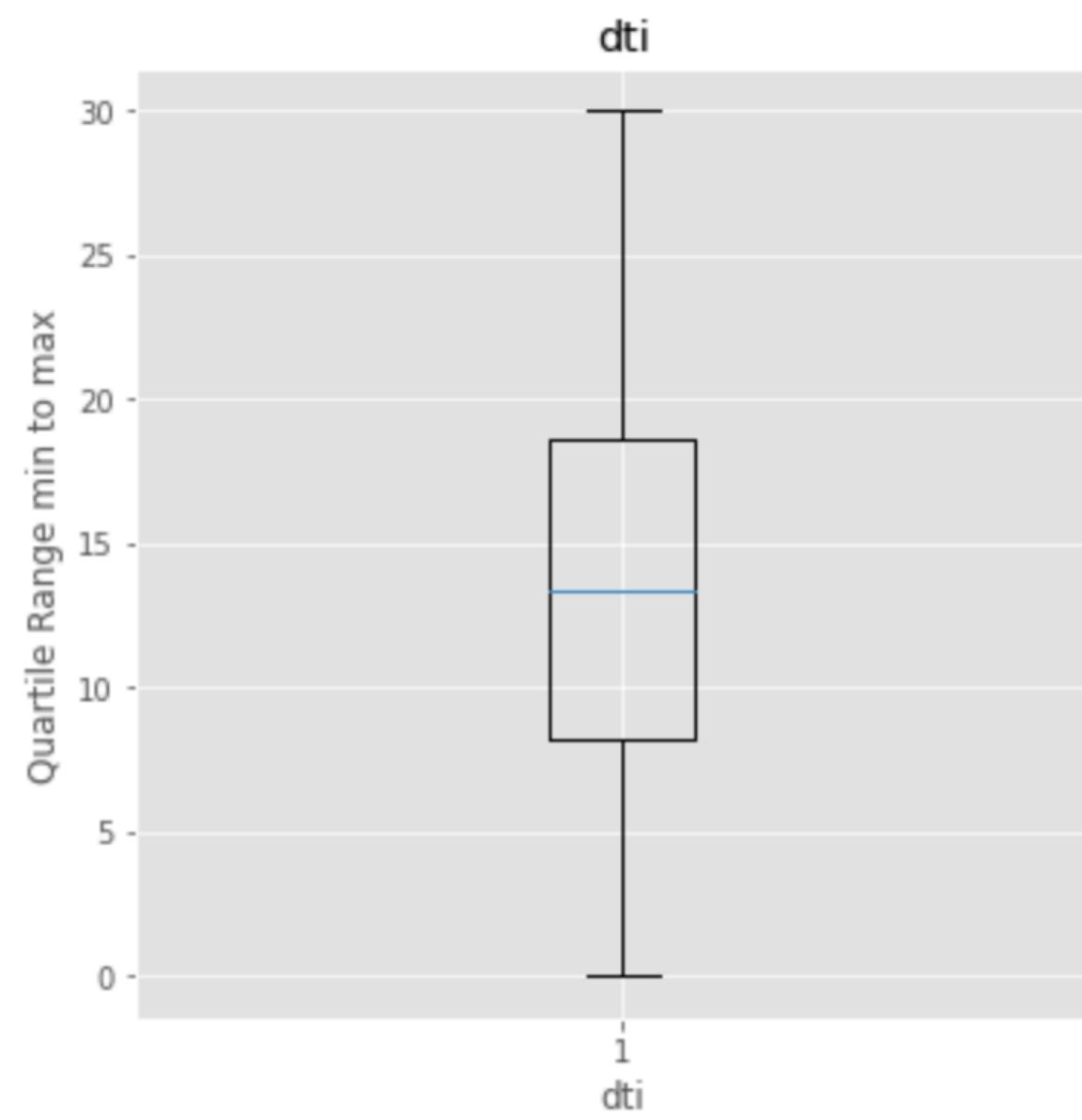
Highlights:

Analysis

Around 25 % of loan application have Debt to income ration greater than Q3 -i.e. 18.5 - These customers can be risky

Around 25 % of loan application have revolving credit utilization greater than Q3- i.e. 72 These Customers can be risky

Lower the Grades higher the chances of Default Can see a continuous increase if default percentage from A to G

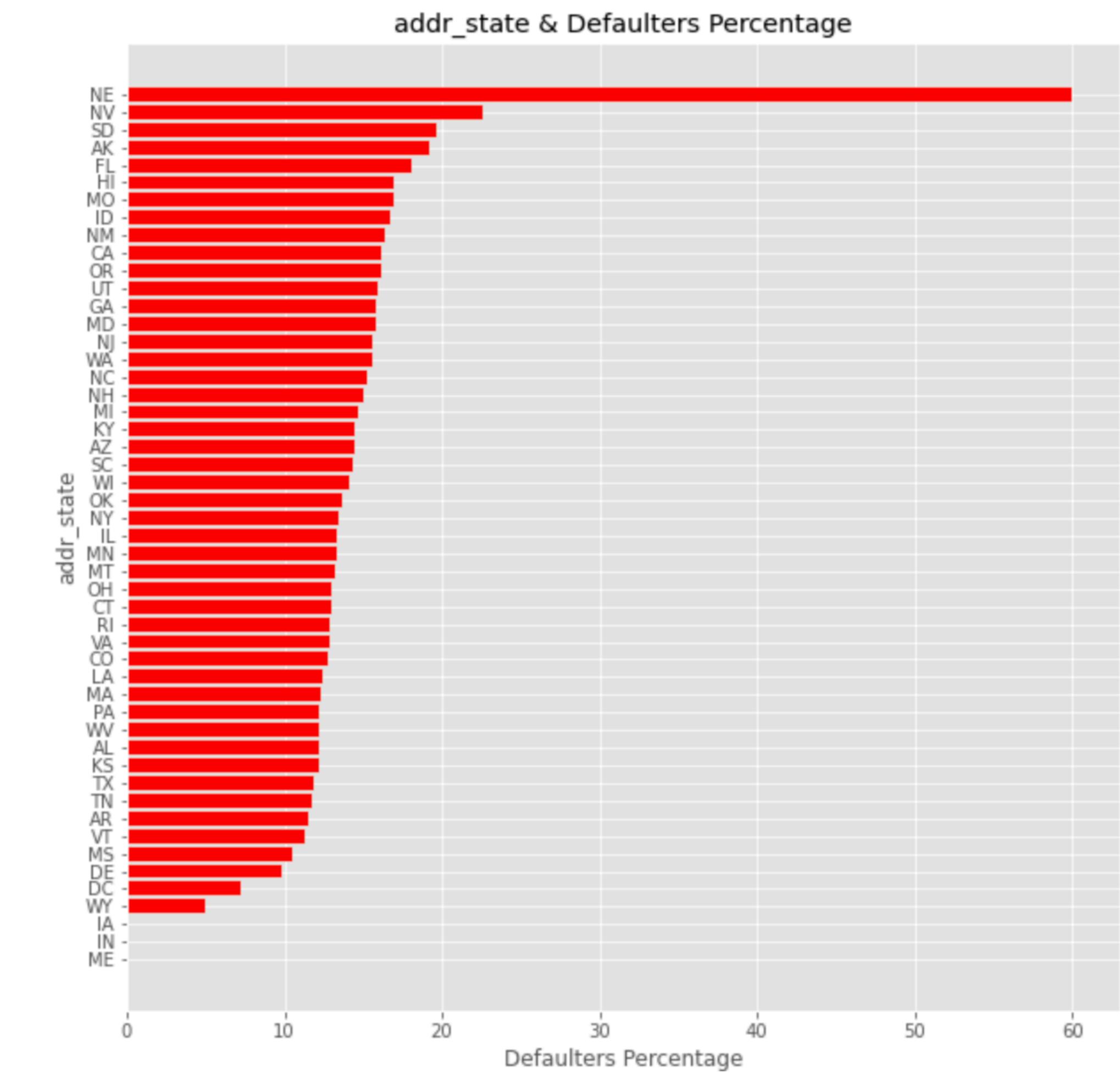
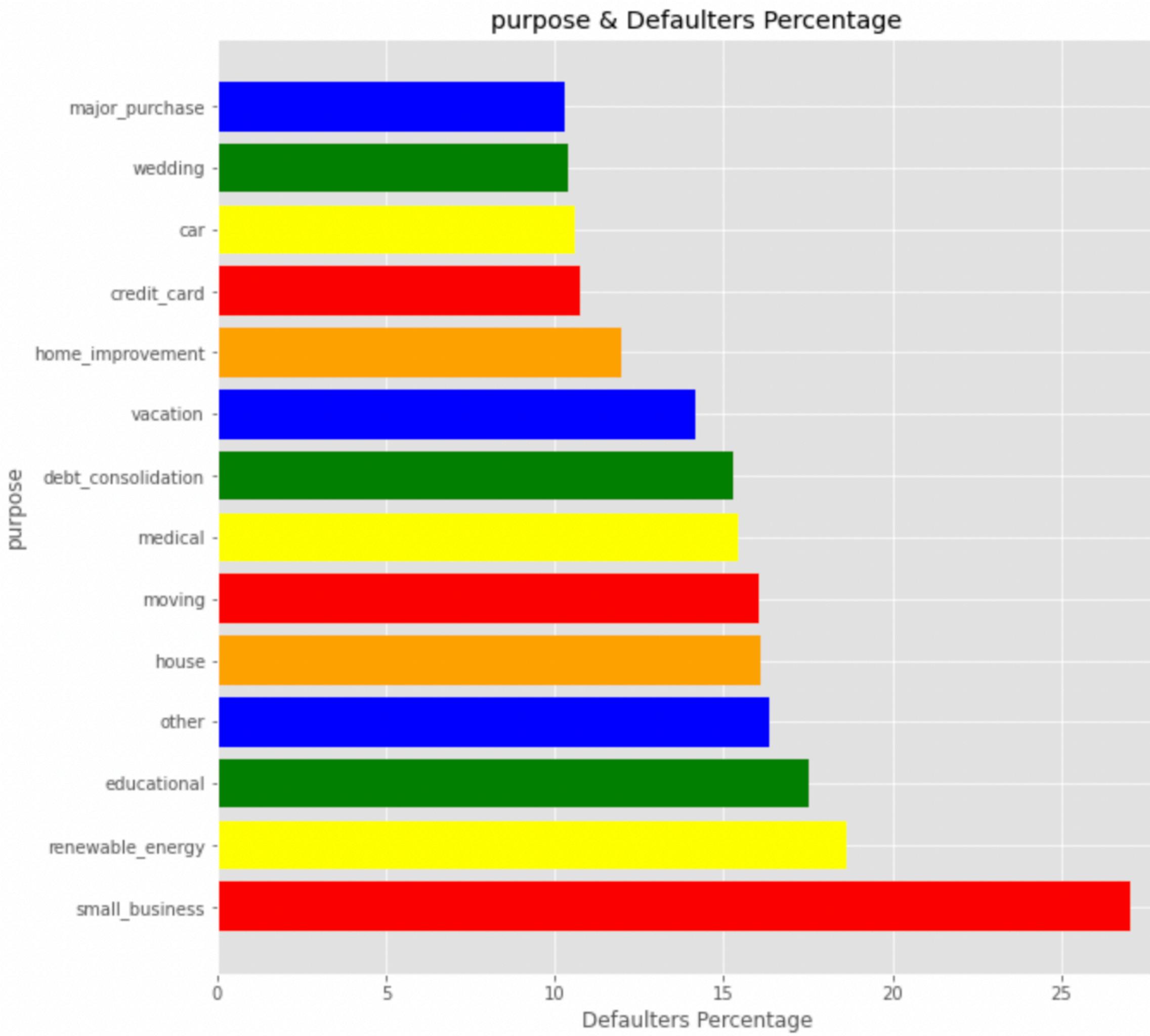




Customers Taking loans for small_business are more likely to be defaulter

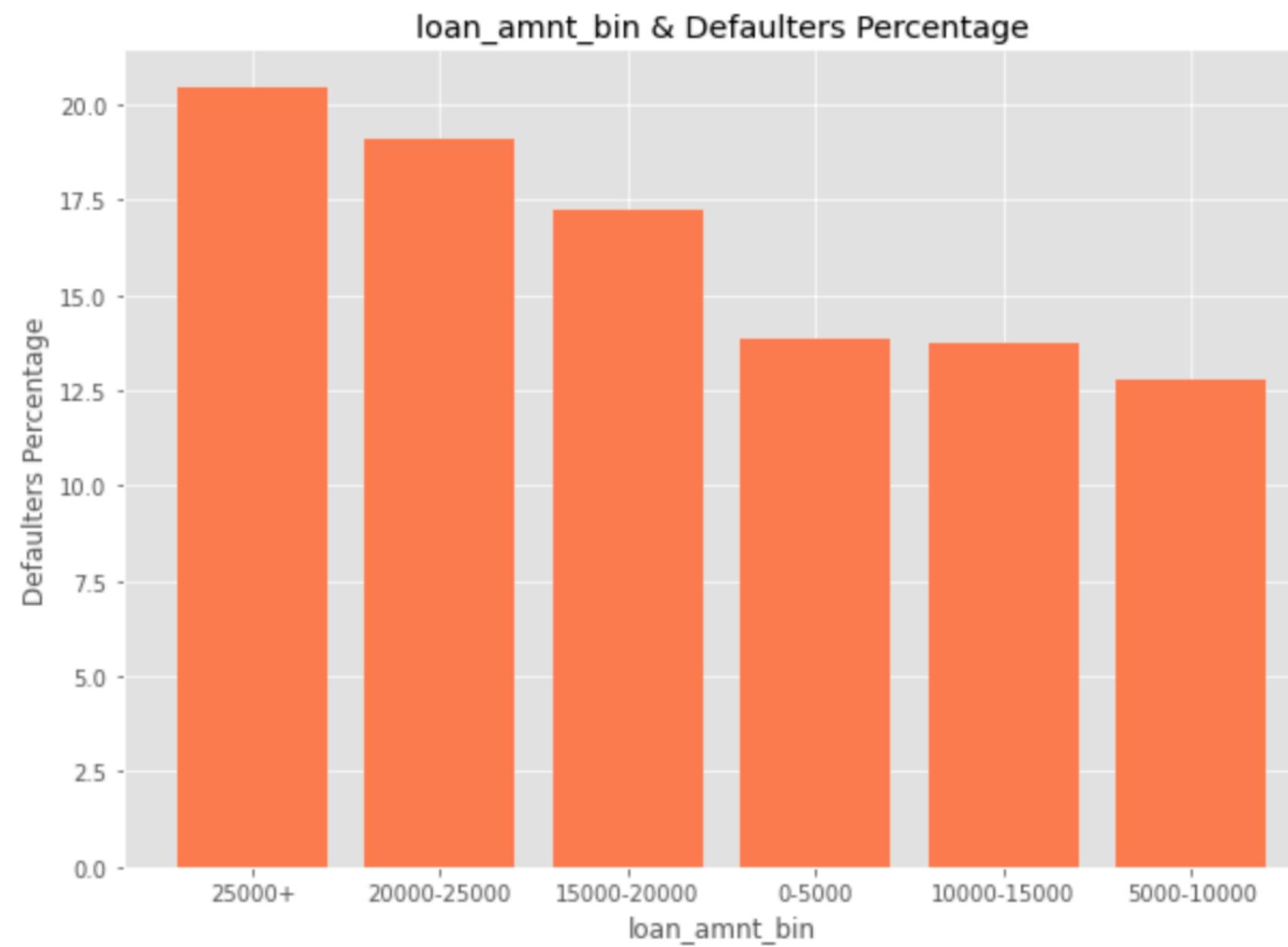
Top 5 states where most no of defaulters are observed can restrict loans for these state NE, NV, SD, FL, HI

Analysis

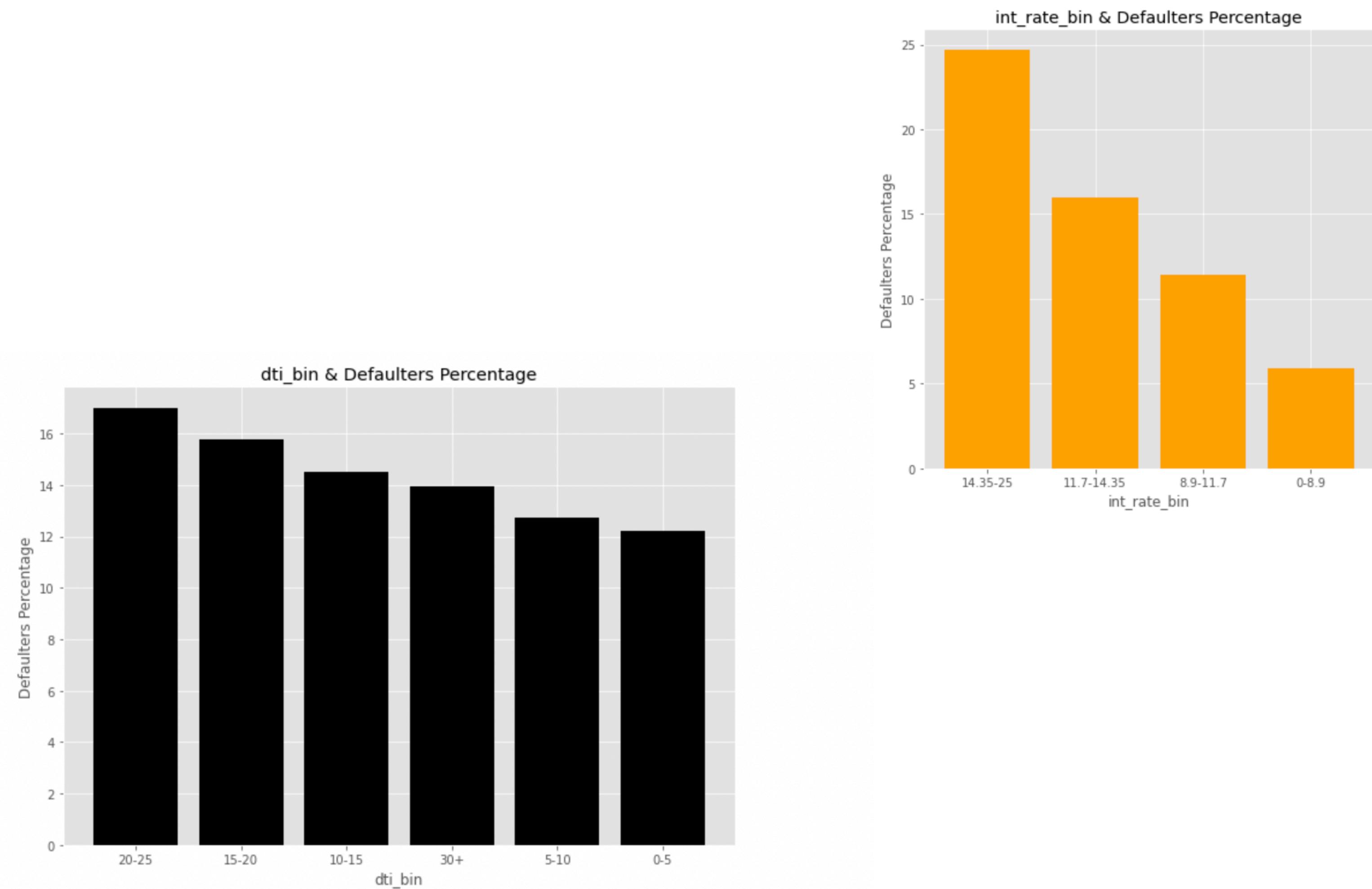




Analysis



Higher the loan amount higher will be the defaulters percentage
Higher the Debt to Investment ratio higher will be the Defaulters Percentage
Higher the Interest rate higher the chances of getting default

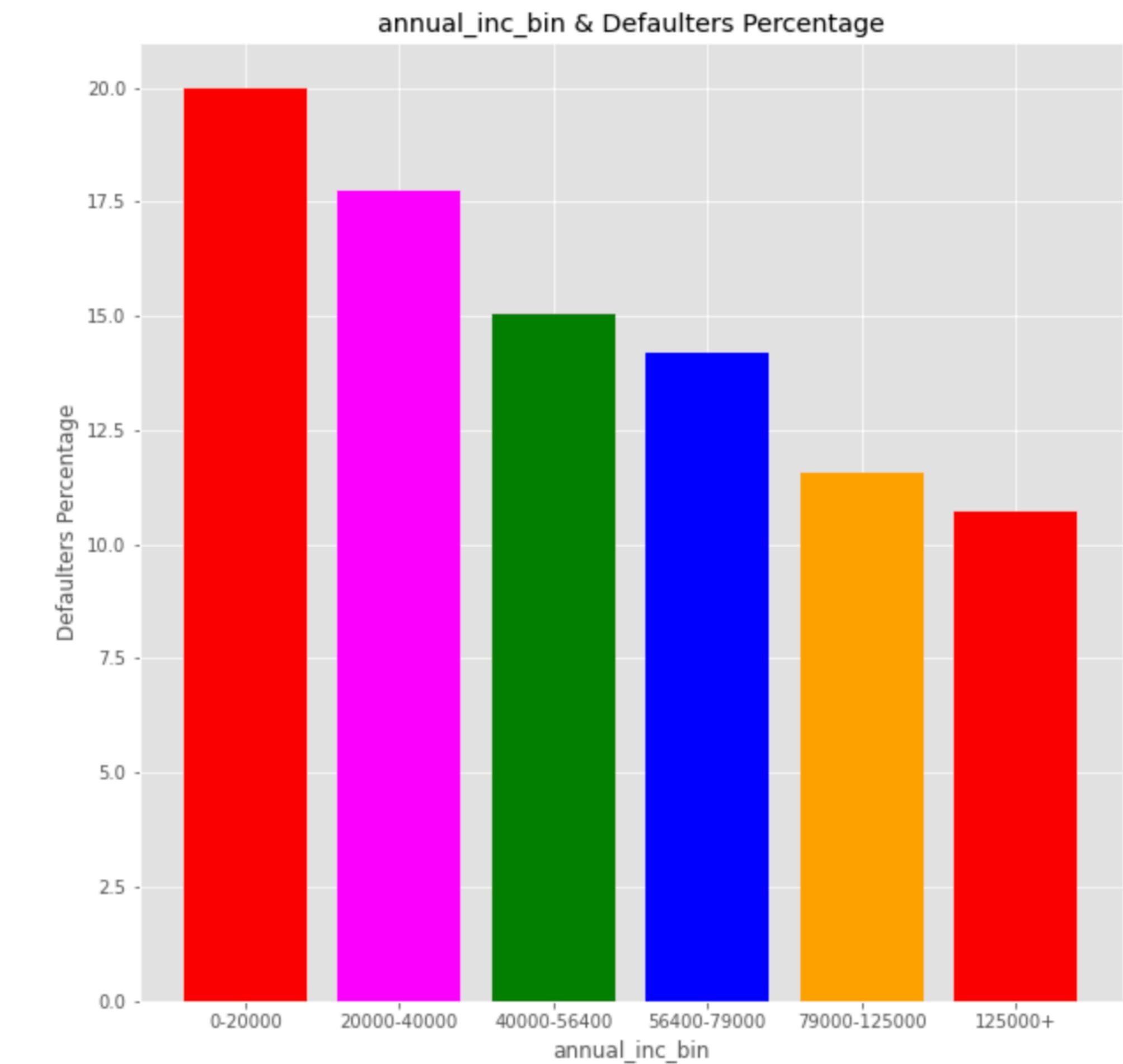
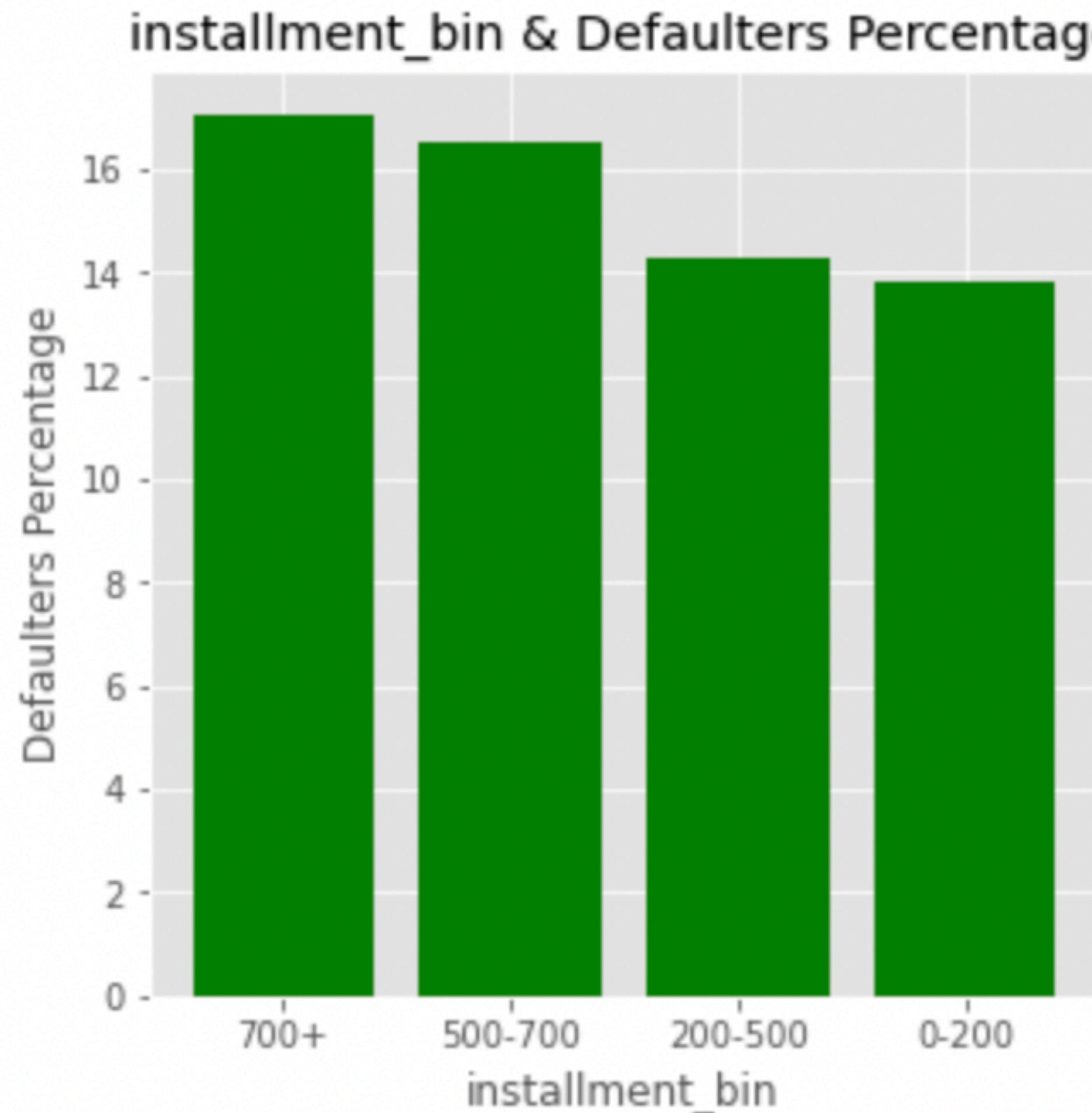




Higher the instalment higher the Defaulter Percentage

Lower the income higher the chances of getting default

Analysis

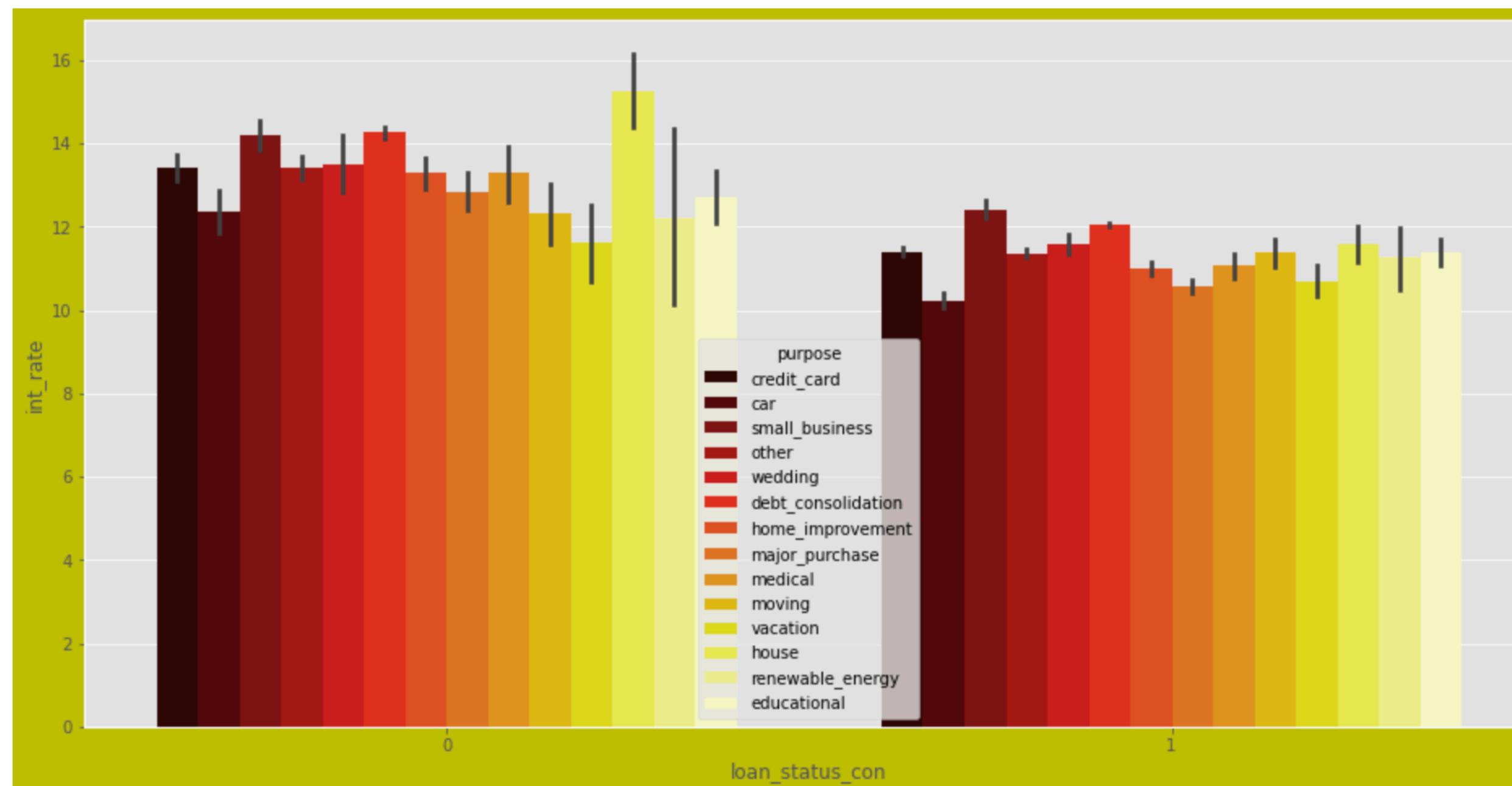


Highlights:

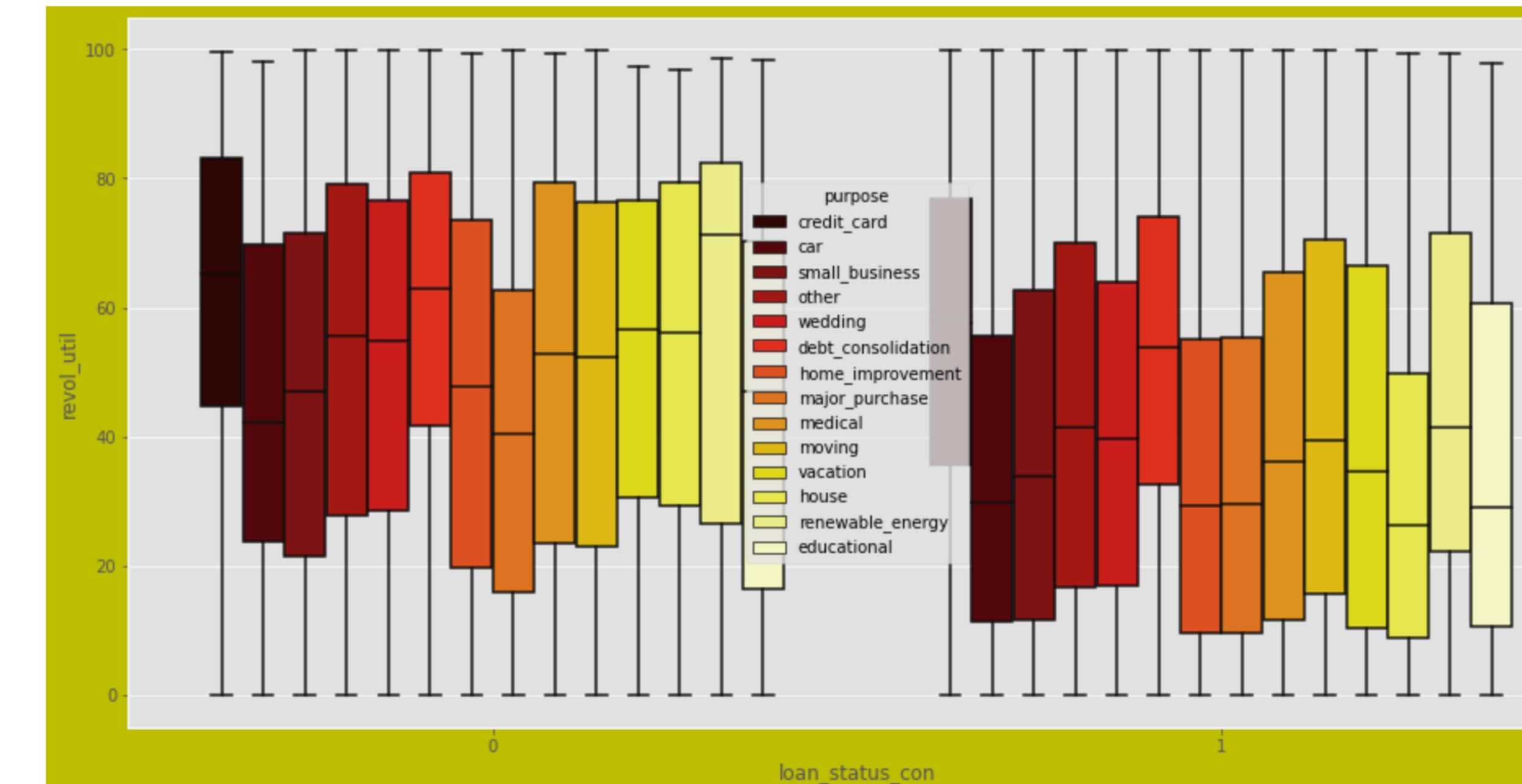
Interest rate for defaulters are high for each purpose
Revol credit utilisation for most of the defaulters are quite high for each purpose

Analysis

Relationship with loan_status_con,int_rate & purpose



Relationship with loan_status_con,revol_util & purpose



0 = “Defaulted”

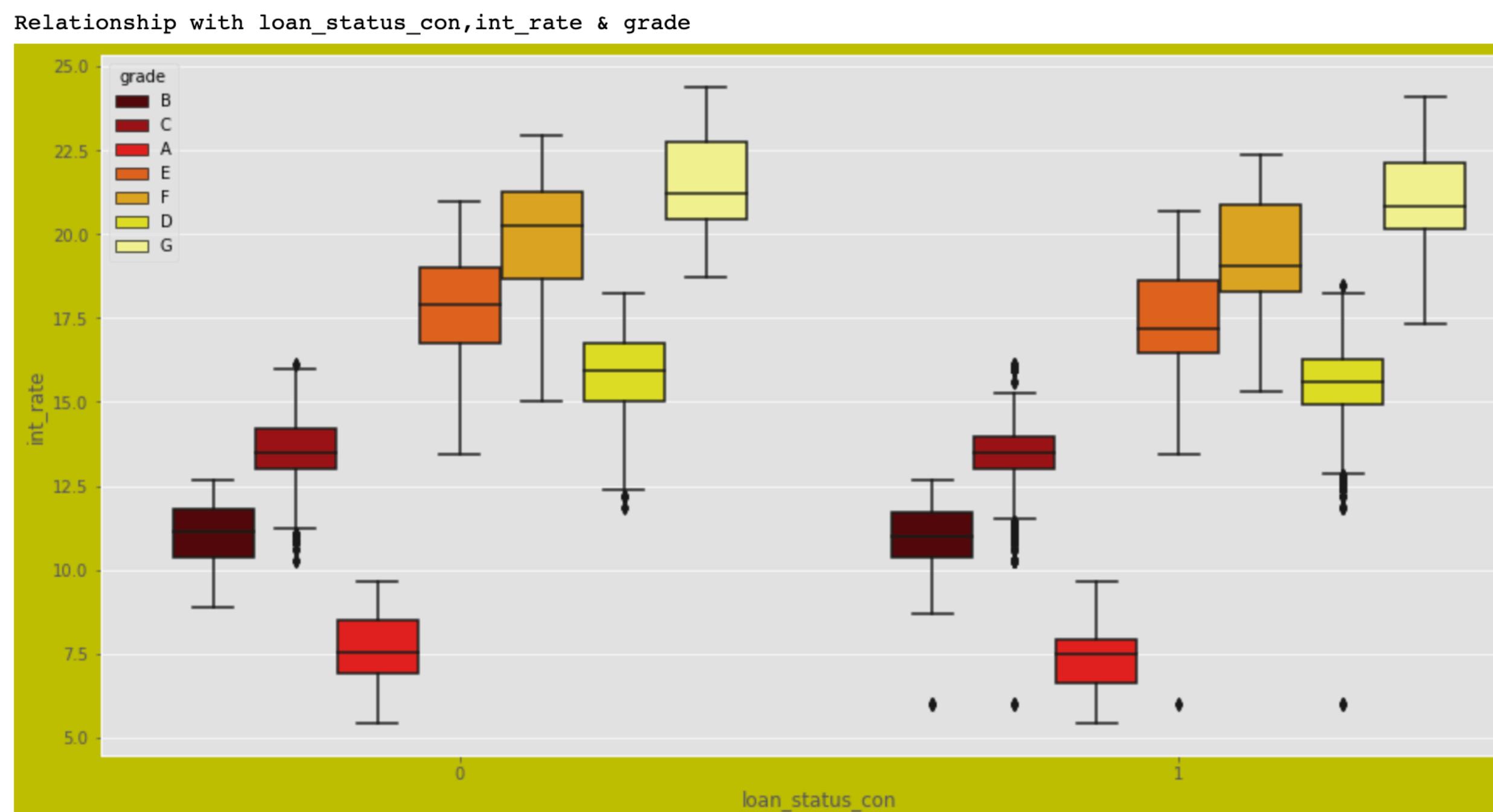
1 = “Fully Paid”

Highlights:

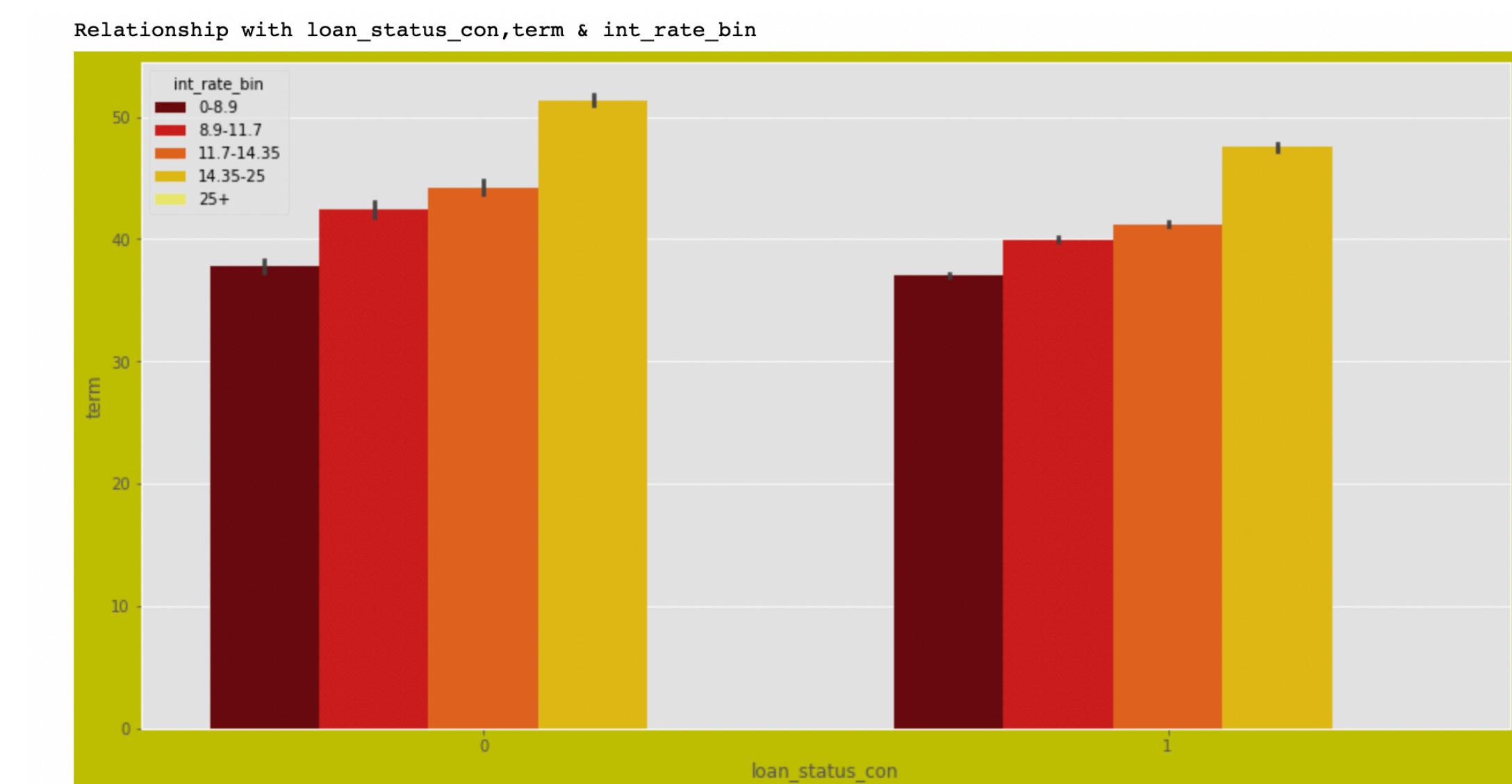
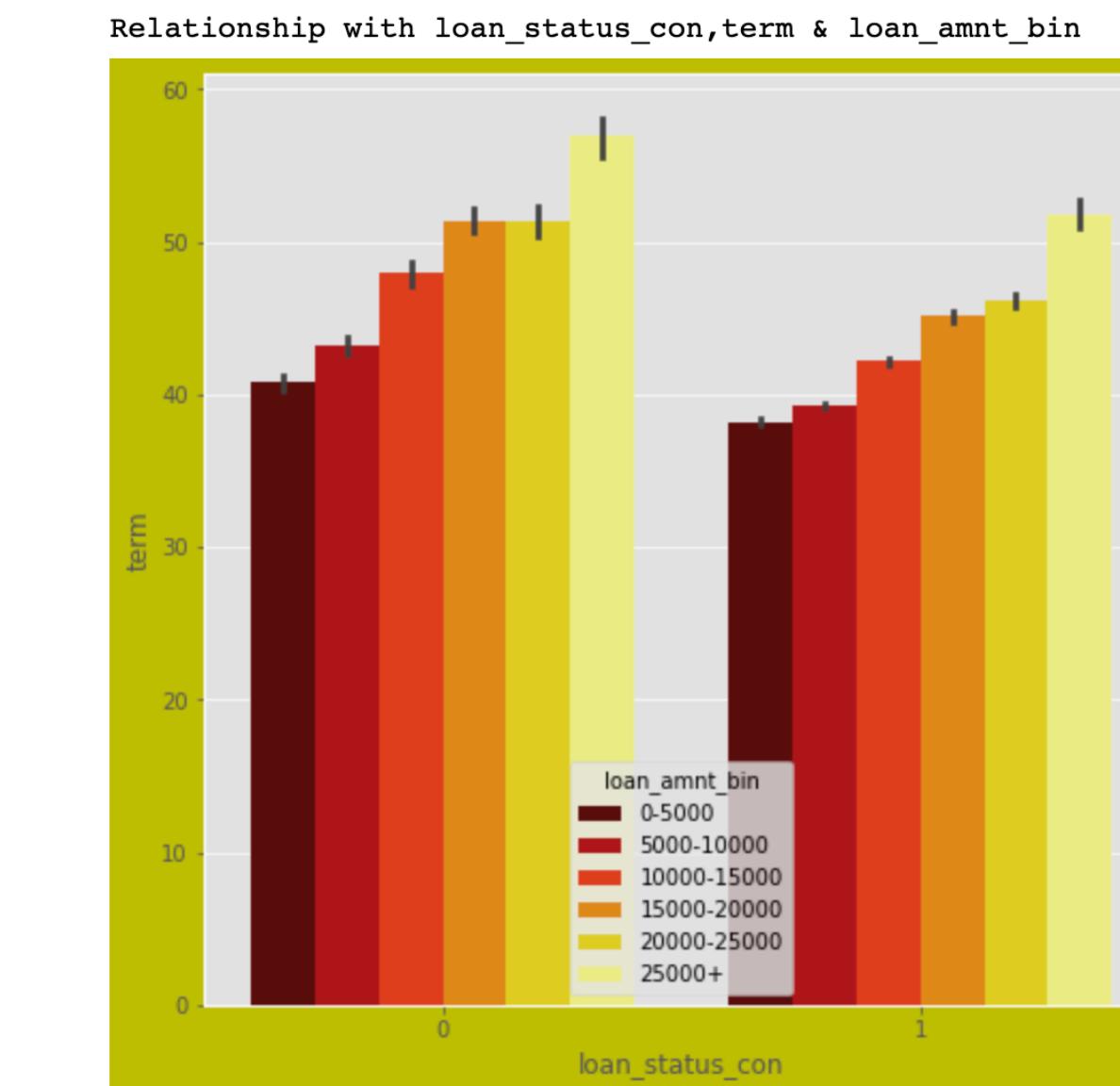
Interest rate for defaulters are high and most of the defaulters of Grade E,F,G,D have their median greater than third Quartile - So we can put a check for approving loan for Grade E,F,G,D for whom the int rate is more than third quartile

Term for defaulters are quite high and higher interest rate for higher rate is most likely to default

Analysis



0 = "Defaulted"
 1 = "Fully Paid"



Highlights:



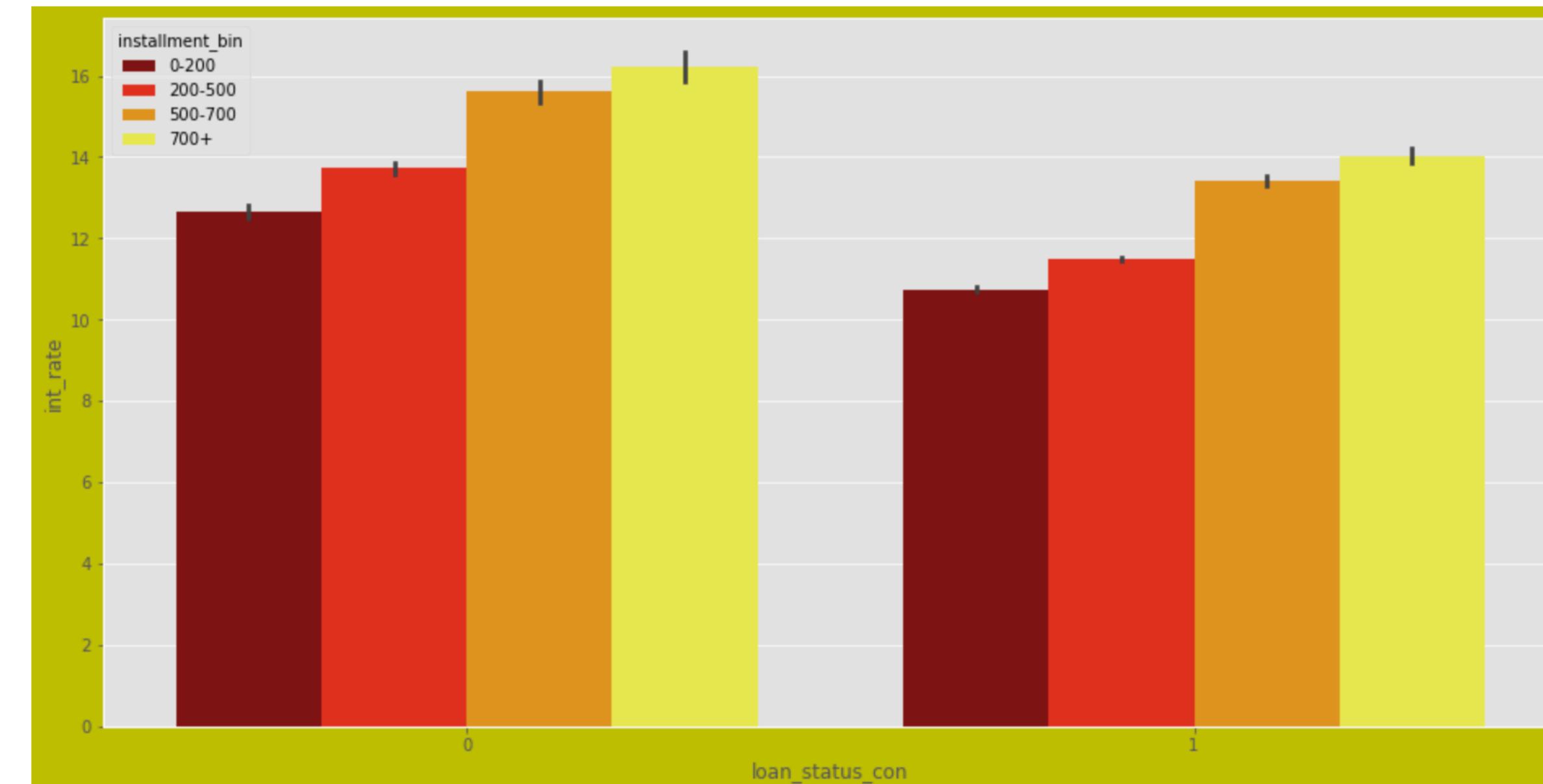
International
Institute of Information
Technology Bangalore

Higher the interest rate and instalment higher the chances of getting Defaulted

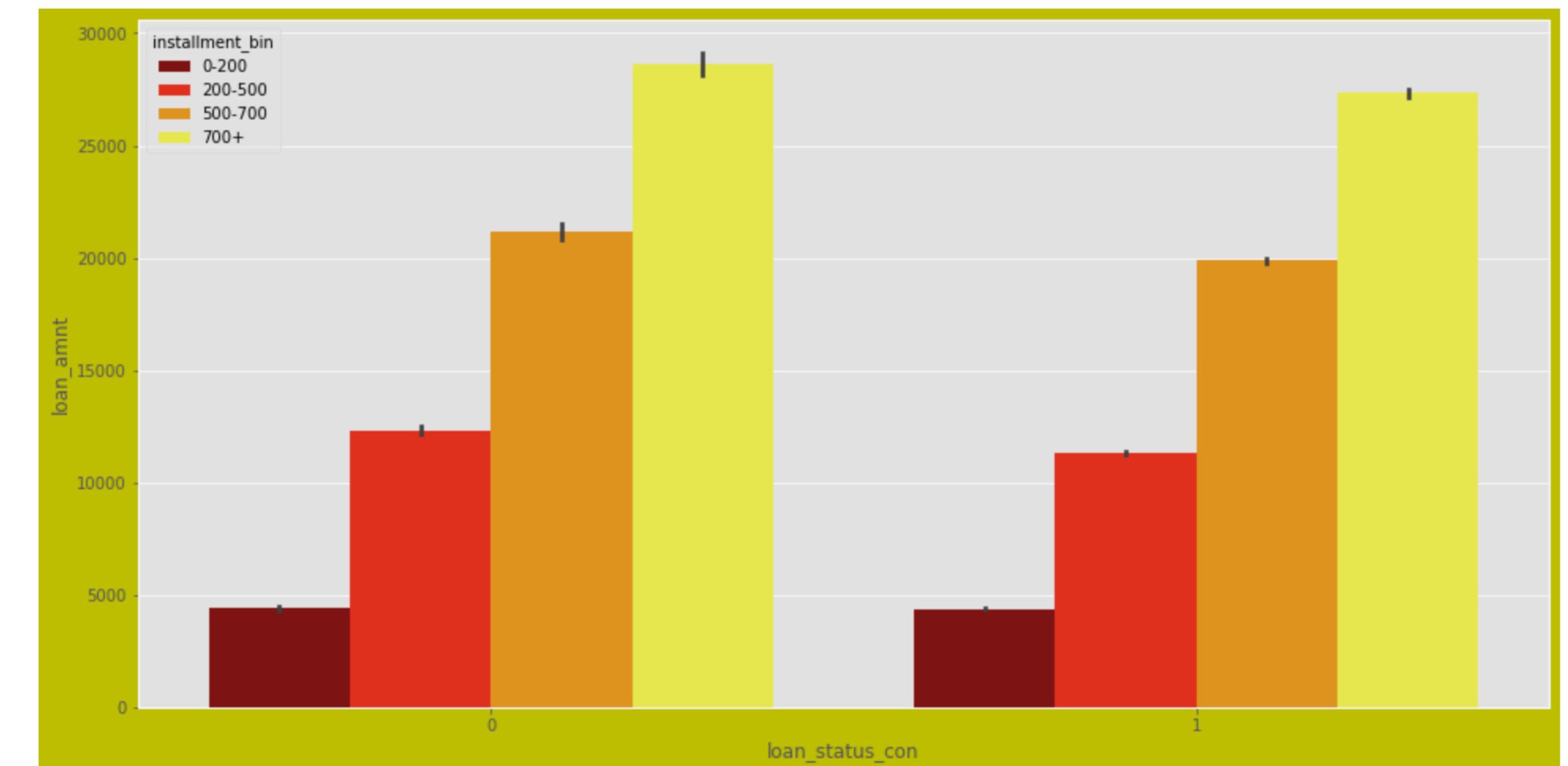
Higher the instalment & loan amount higher the chances of getting Defaulted

Analysis

Relationship with loan_status_con, int_rate & installment_bin



Relationship with loan_status_con, loan_amnt & installment_bin



0 = “Defaulted”
1 = “Fully Paid”

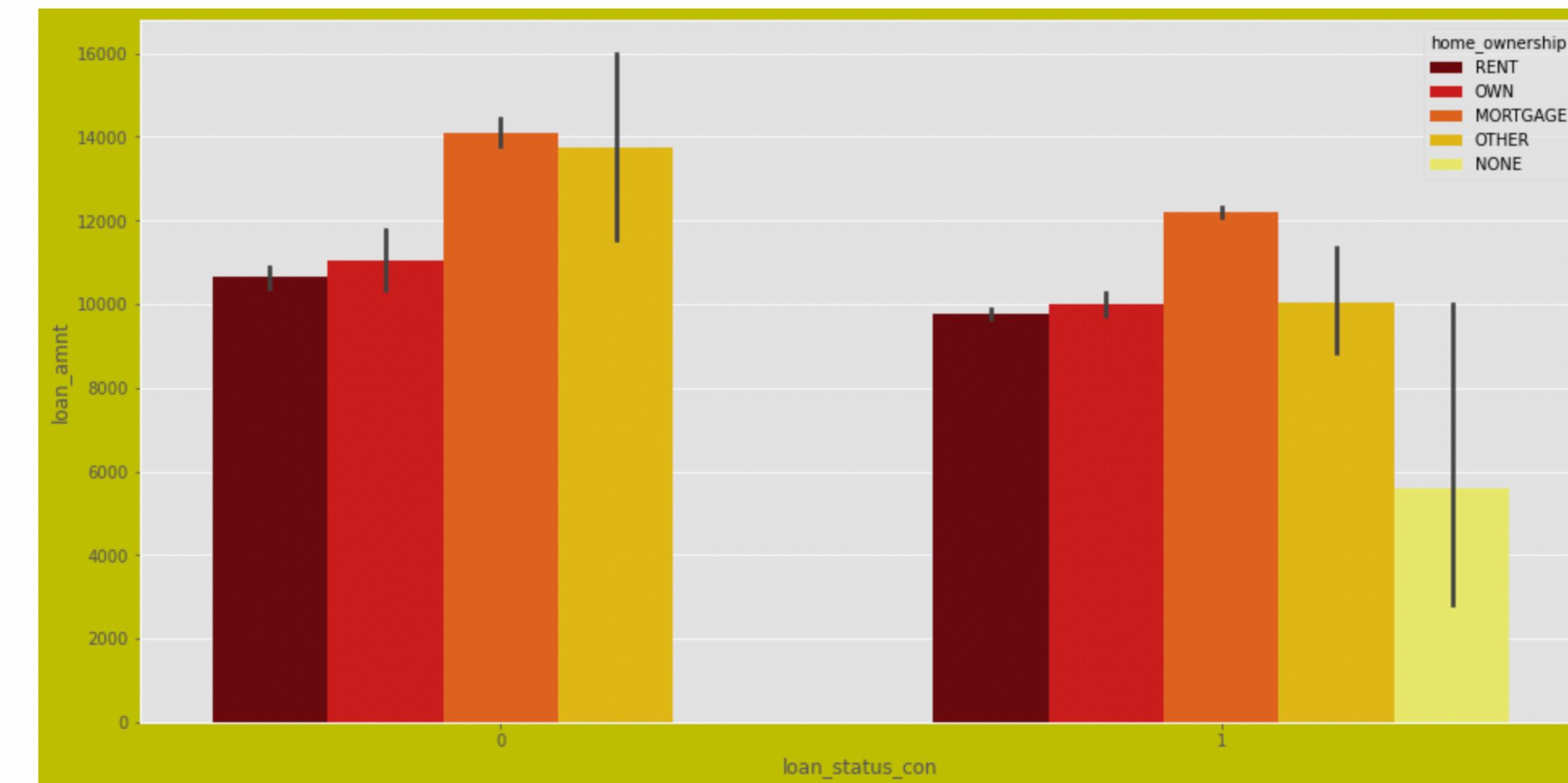
Highlights:

Higher Defaulters are present for home type Mortgage with respect to high loan amount

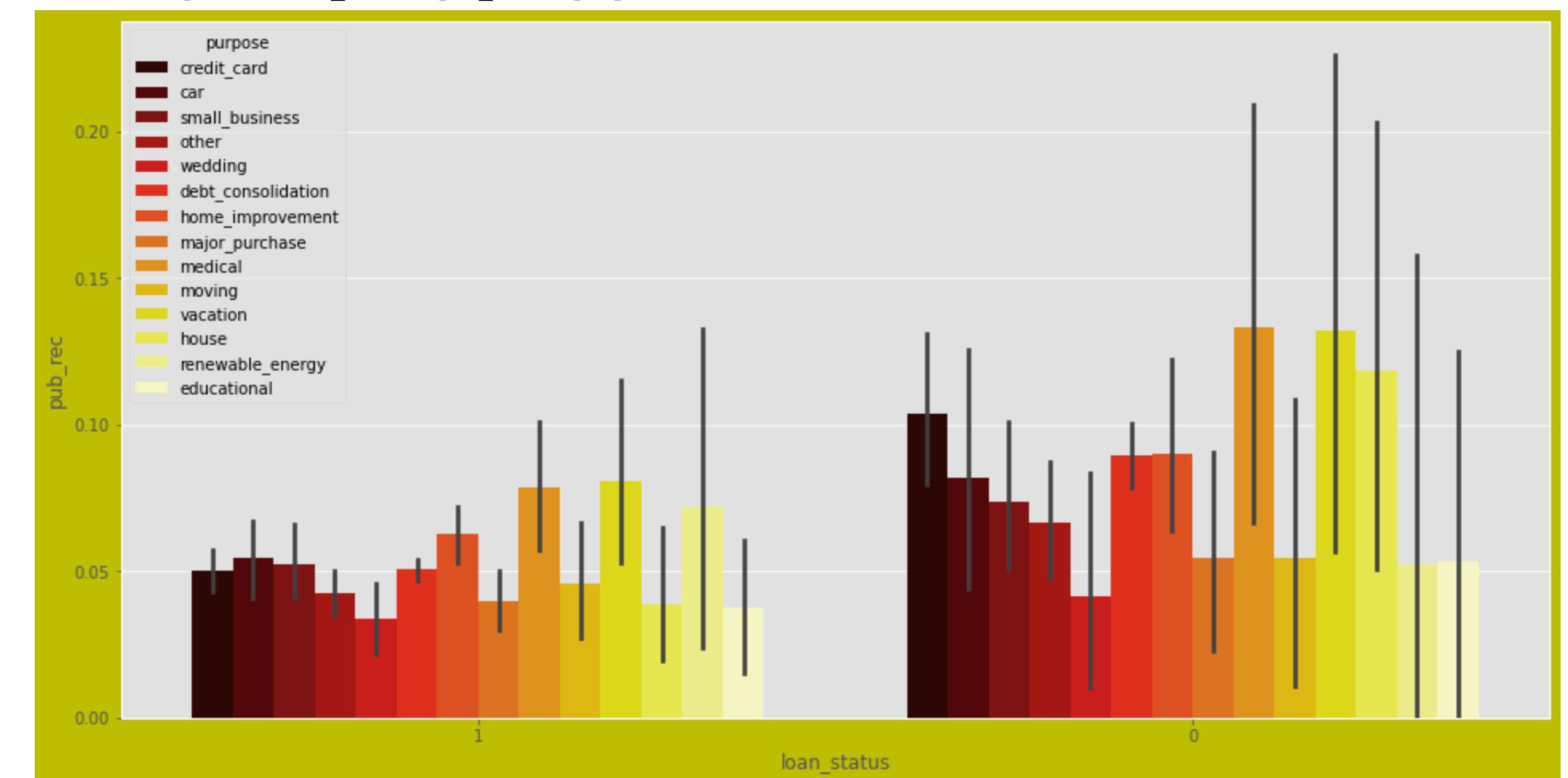
Higher no of public records higher the chances of getting defaulted

Analysis

Relationship with loan_status_con, loan_amnt & home_ownership



Relationship with loan_status, pub_rec & purpose



Conclusion

- Loan Applications must be verified
- Lower the Grades higher the chances of getting default , Before approving loans to lower grades it should be thoroughly investigated
- small_business loans are most likely to be defaulted lending institution should reduce the no of loans issuing to this category
- Higher the loan amount , interest , installment and debit to investment ratio higher the chances of getting default , Lending club should pay attention to these attributed before allocating the loans
- Lending club should keep a check on high revolving credit utilisation before approving the loan
- Lending club should try to try reduce the term higher the term higher the chances of getting default
- Public record borrower are most likely to bankrupt lending club should investigate before approving loan that there are no public records for borrower
- Lending club should reduce the no of loans to borrowers with home type Mortgage with higher loan amount as their chances of getting defaulted is more