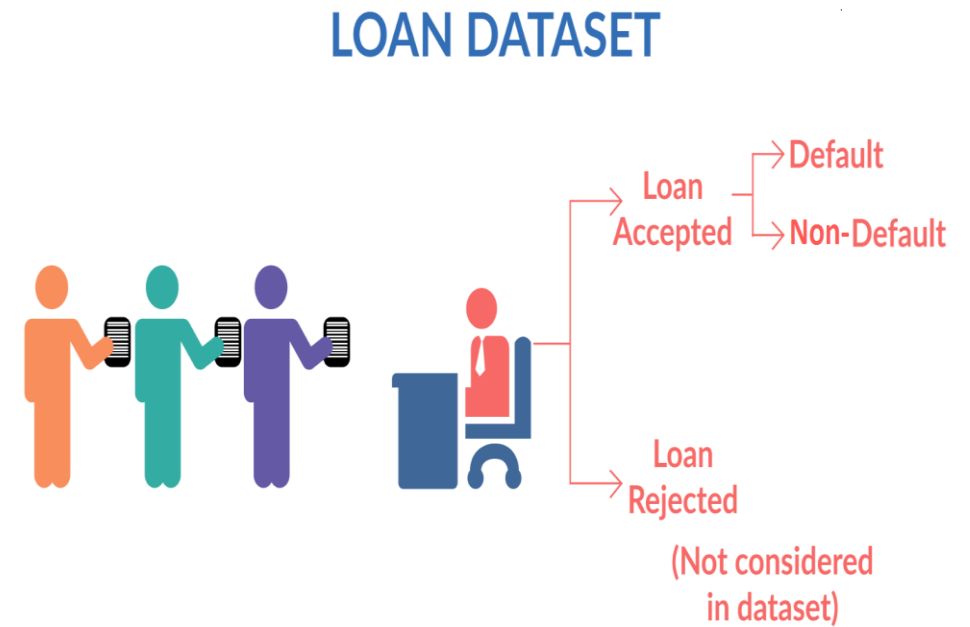# LENDING CLUB CASE STUDY

SUBMISSION

Name: Deepak Gadhamsetty; K S Gopalakrishnan

Lending Club – Business

- Online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures.

- lending loans to 'risky' applicants is the largest source of financial loss when the borrower refuses to pay or runs away with the money owed.

- A loan application is accepted or rejected.

- Dataset has records of loans accepted with information including status of loan repayment.

LOAN DATASET

Loan Accepted → Default
Loan Accepted → Non-Default

Loan Rejected
(Not considered in dataset)

Lending Club – Problem Statement

The company wants to understand the **driving factors (or driver variables)** behind loan default, i.e. the variables which are strong indicators of default.

Approach to solving the problem

- The columns include variables related to borrower characteristics, loan characteristics, Lending Club created variables (grade, sub grade) and information after approving the loan. For our problem we only need relevant variables from loan characteristics and borrower characteristics available before a approve/reject decision is taken on the loan application. In other words, other variables are not required.

- Within the borrower characteristics and loan characteristics there are certain variables providing similar information. Some variables do not have standardized information. Examples – State and zip code; description of purpose for taking loan, employee title. These have been excluded for further analysis.

- Loan status has fully paid, charged off and current. As we are interested in default rate, records with "current" loan status can be removed.
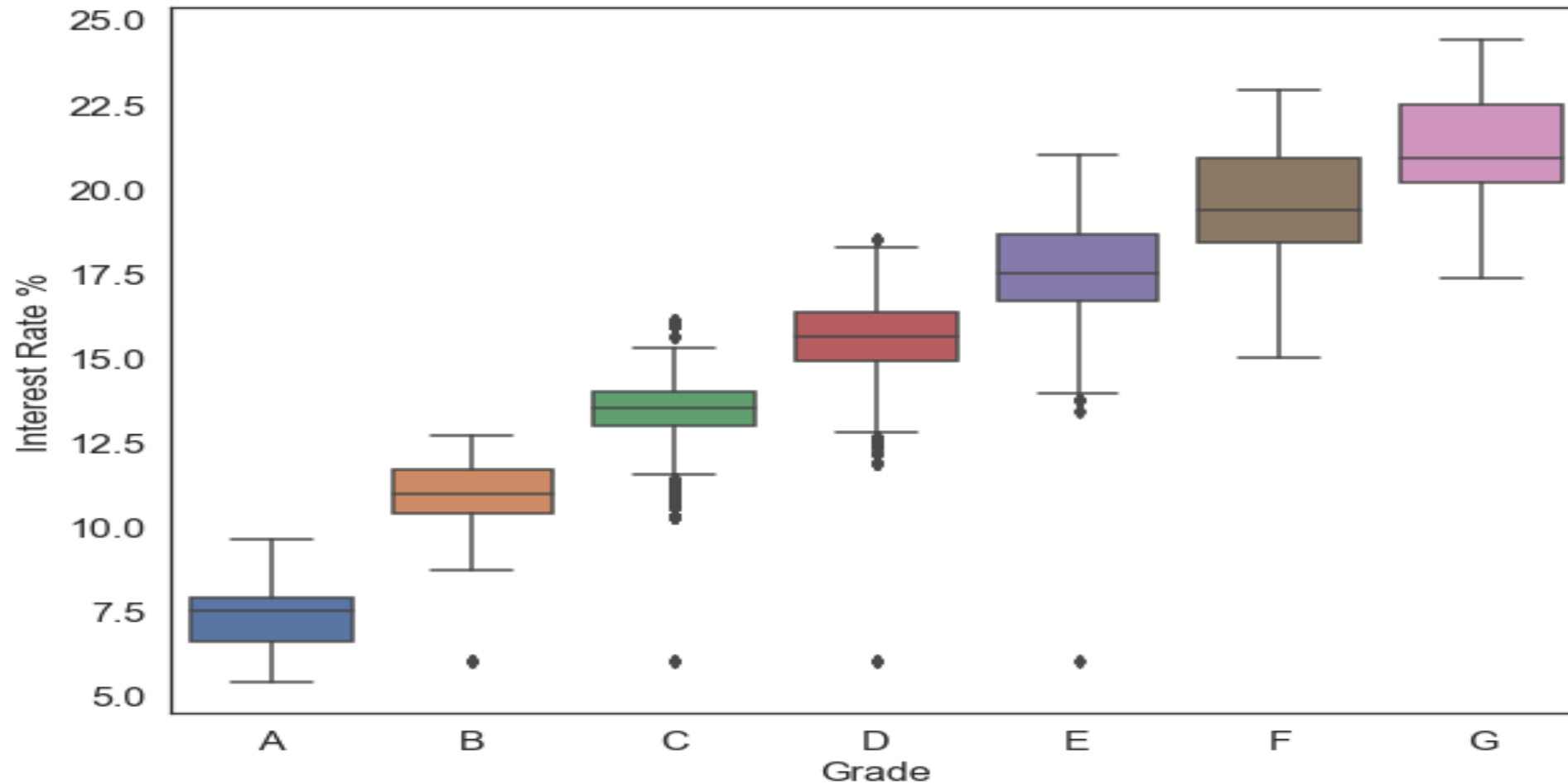
Data Summary

- Data has 39717 records of 111 columns.
- Cleaned up data for null values, outliers, duplicates and errors.
- Final data 38577 records

Data Analysis

14.6% loans default; new business growing steeply year after year
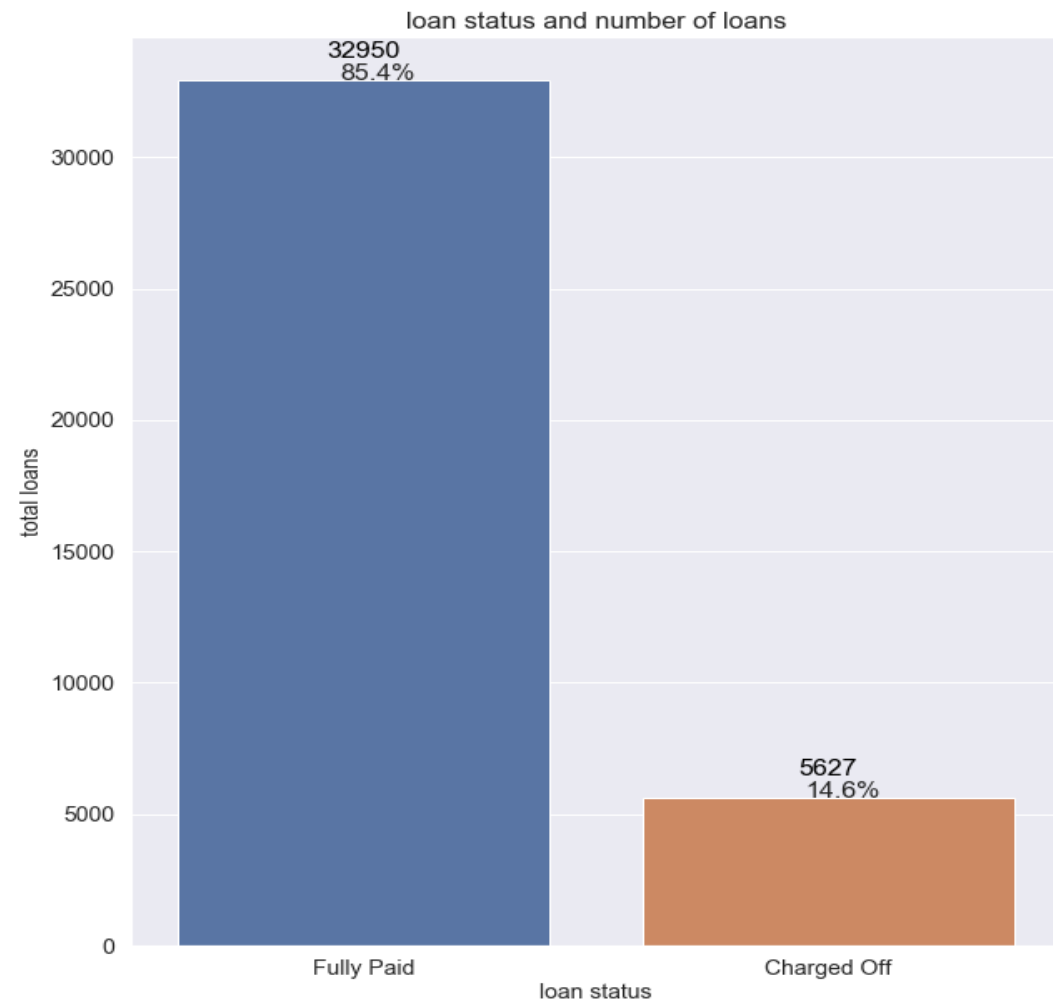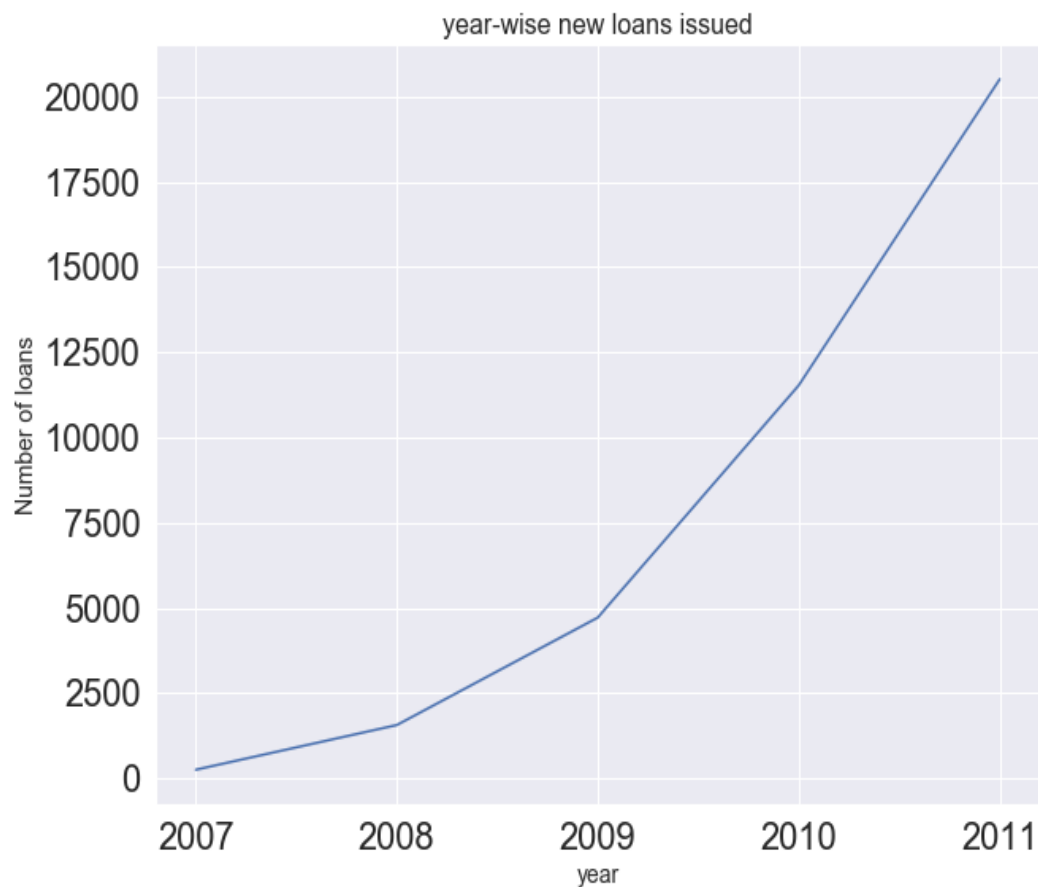
Interest rate increases as grades move from A to G

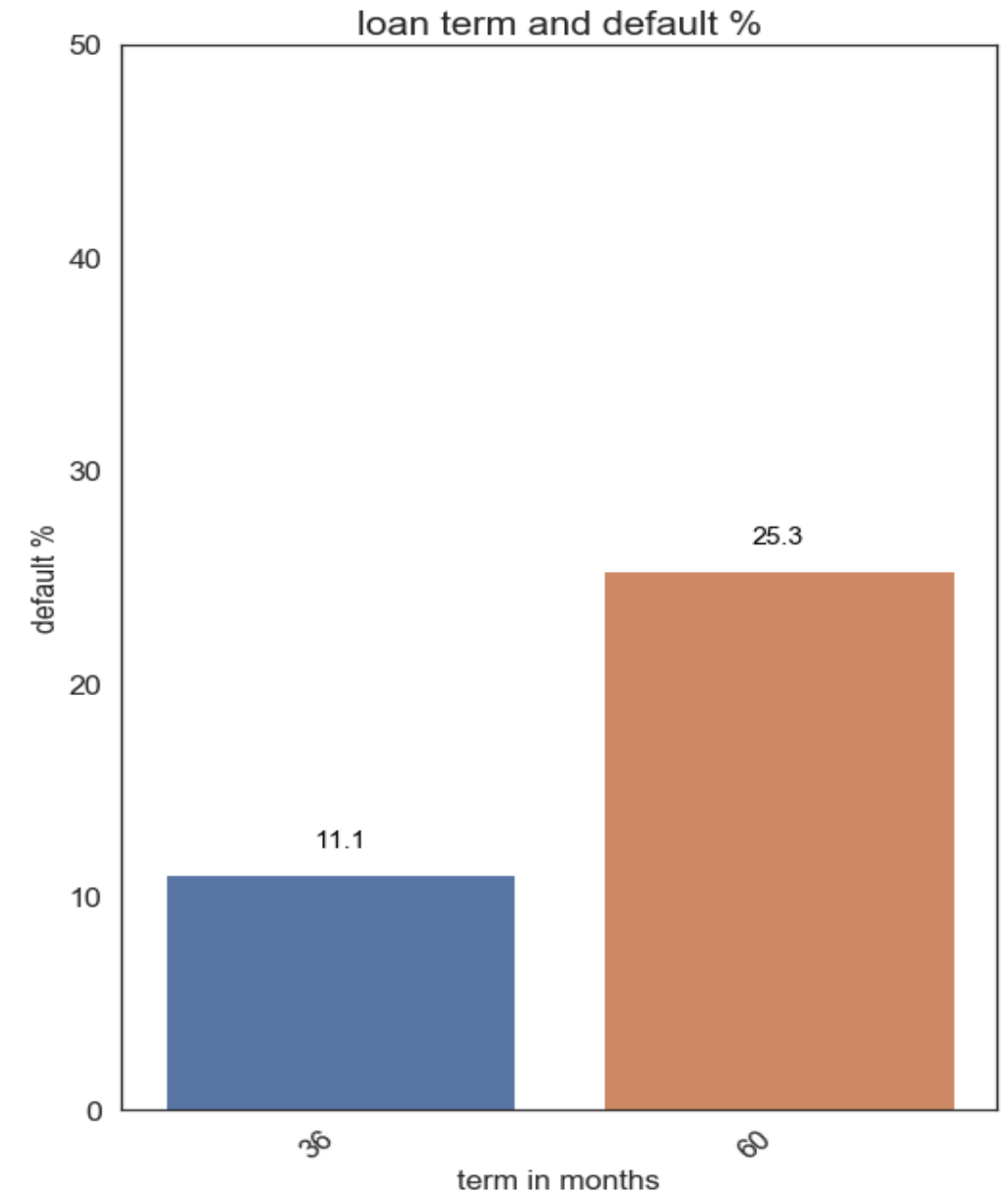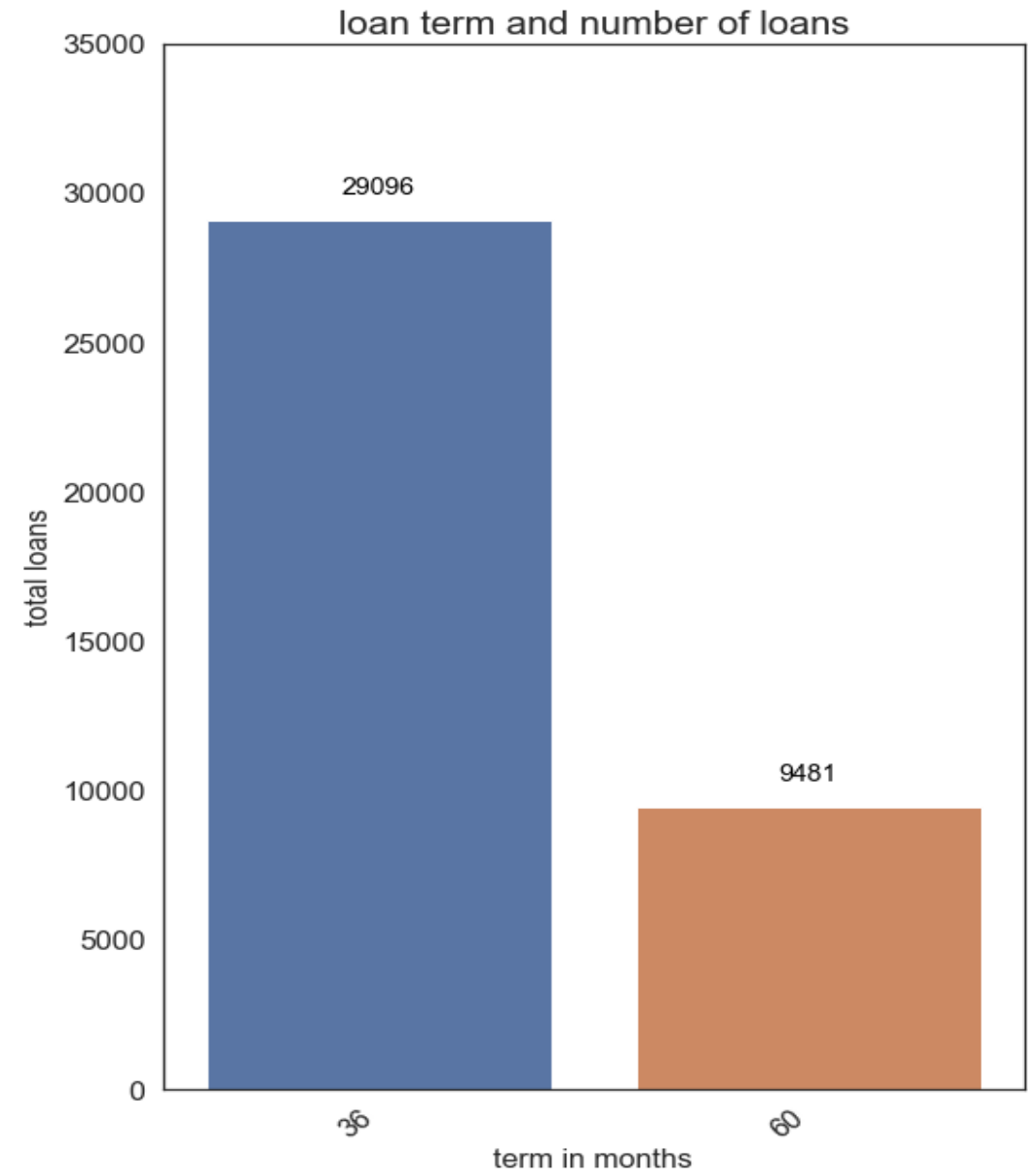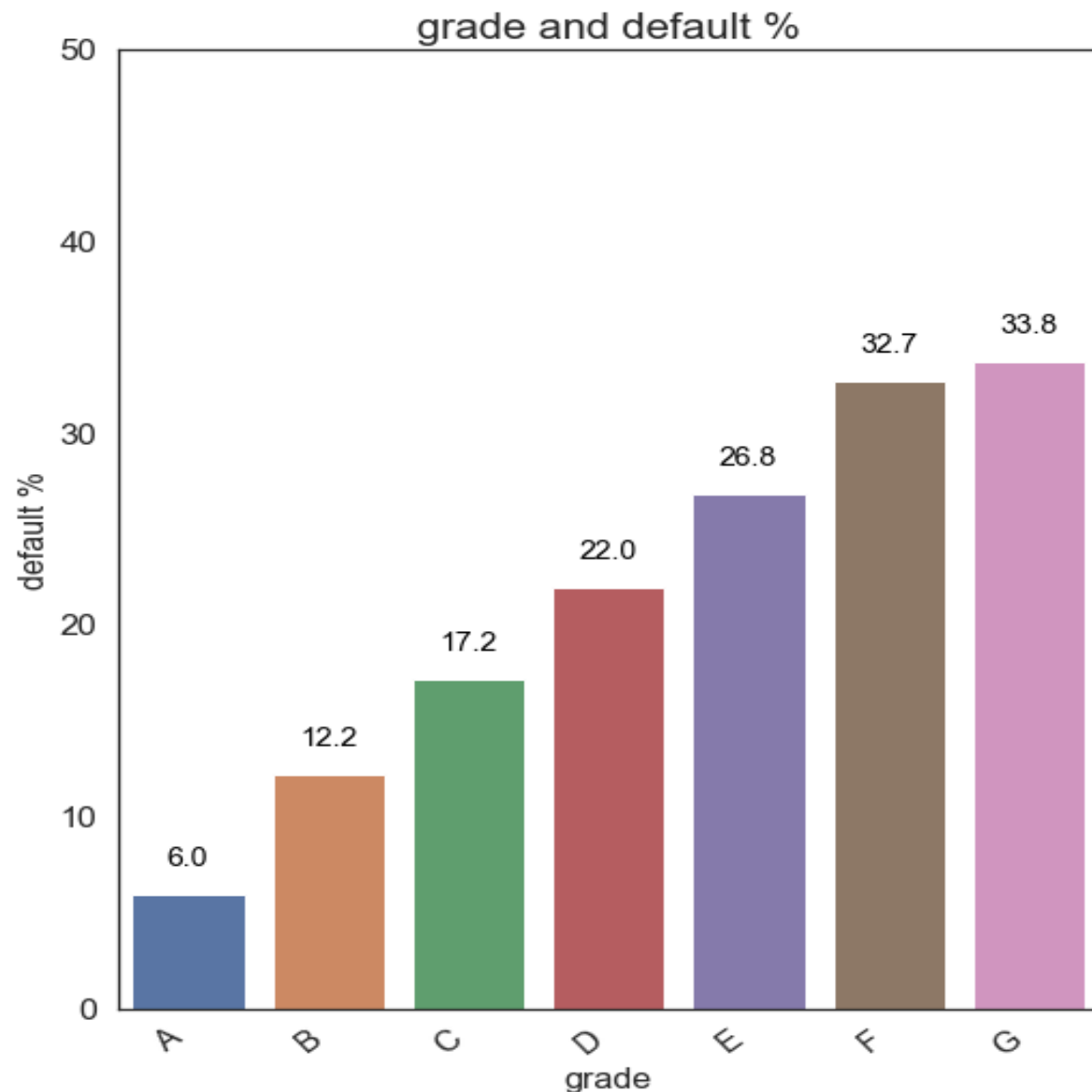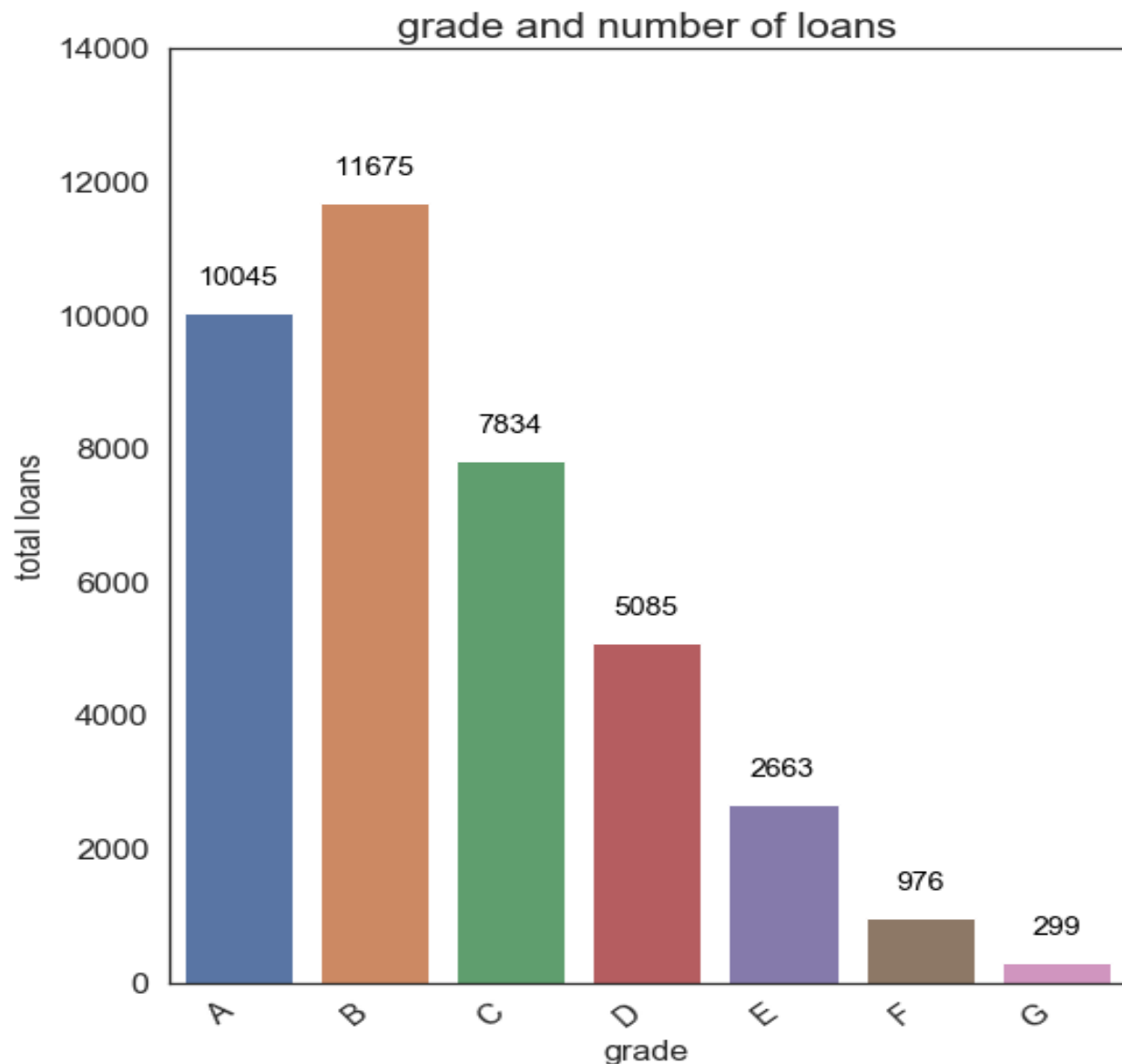Implies A has lowest default risk and G has the highest

Univariate Analysis

Looking for any interesting patterns between default rate and each of the chosen variables individually

Default rate higher for 36 months though number of applications is lower
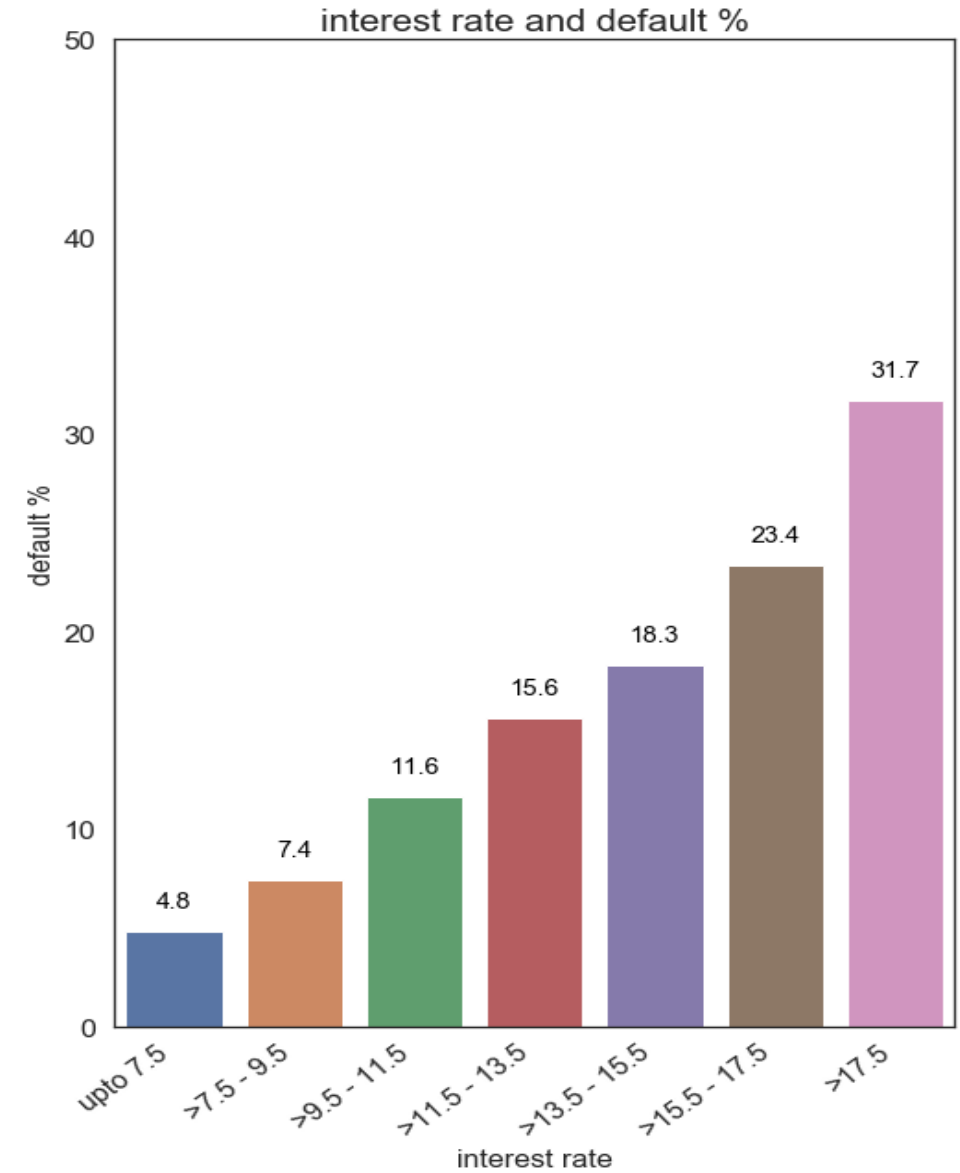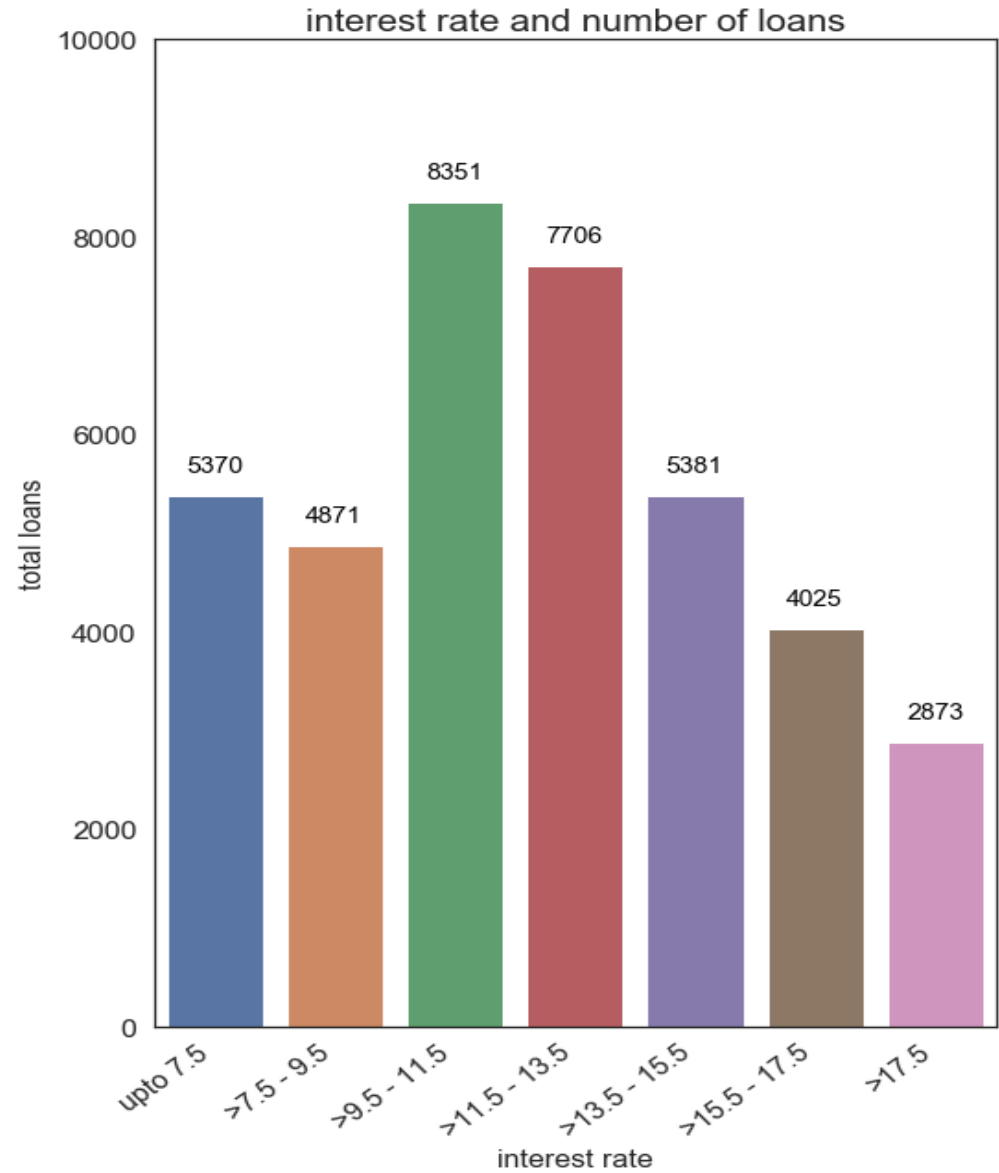
Default rate increases as grade moves from A to G, though number of applications is lower for grades E, F, G
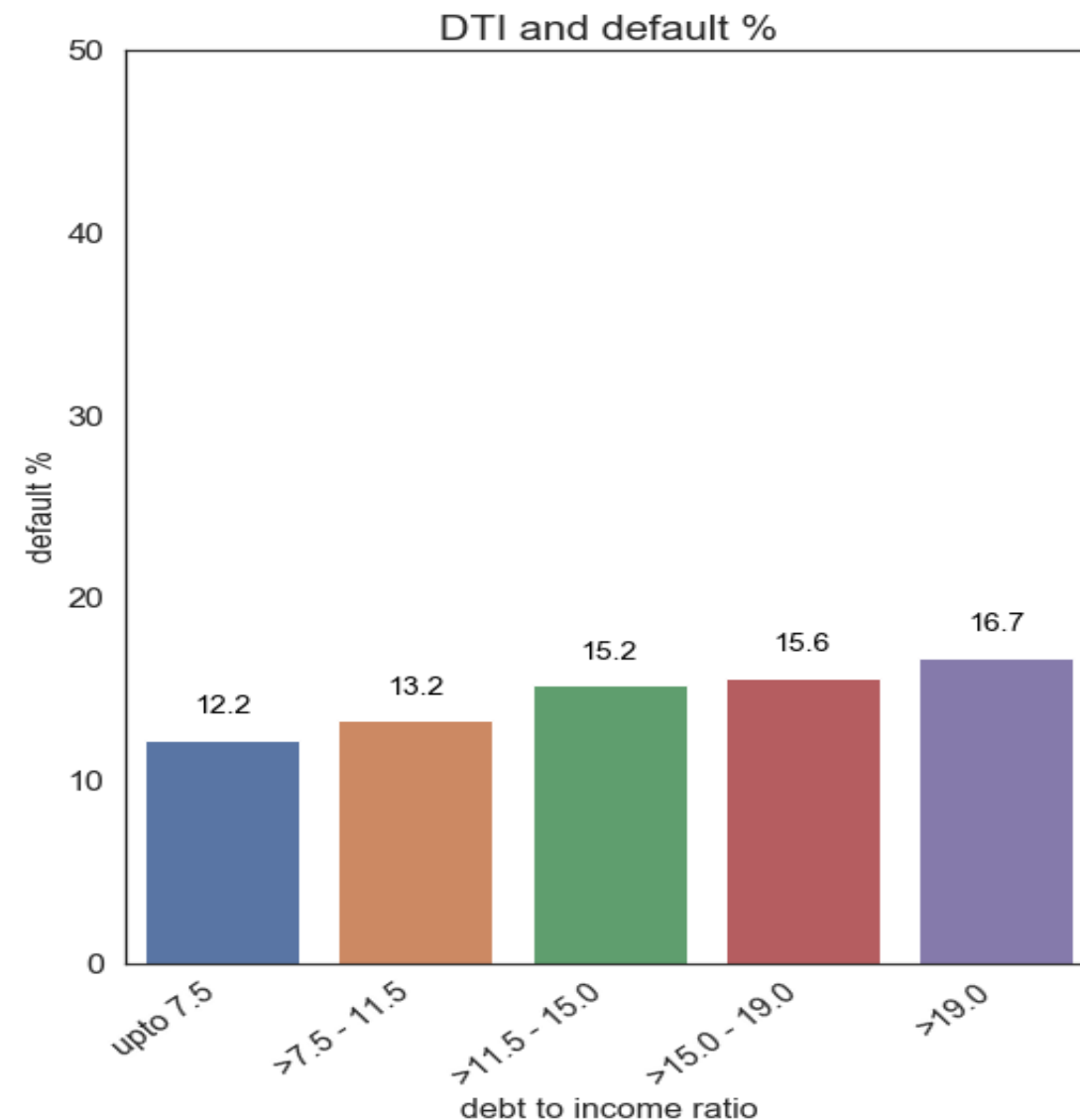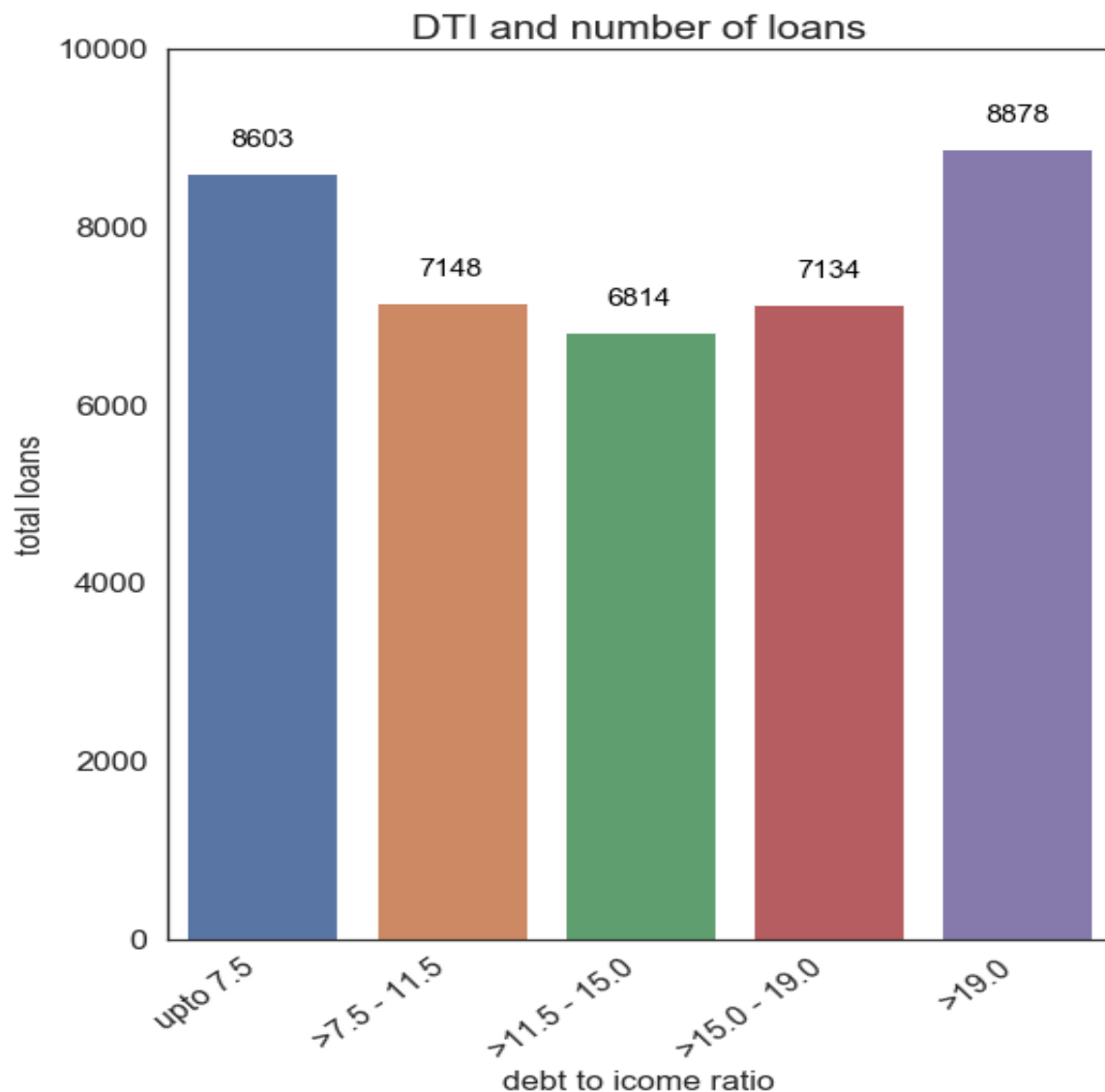
Default rate increases as interest rate increases. Consistent with trends for grades.
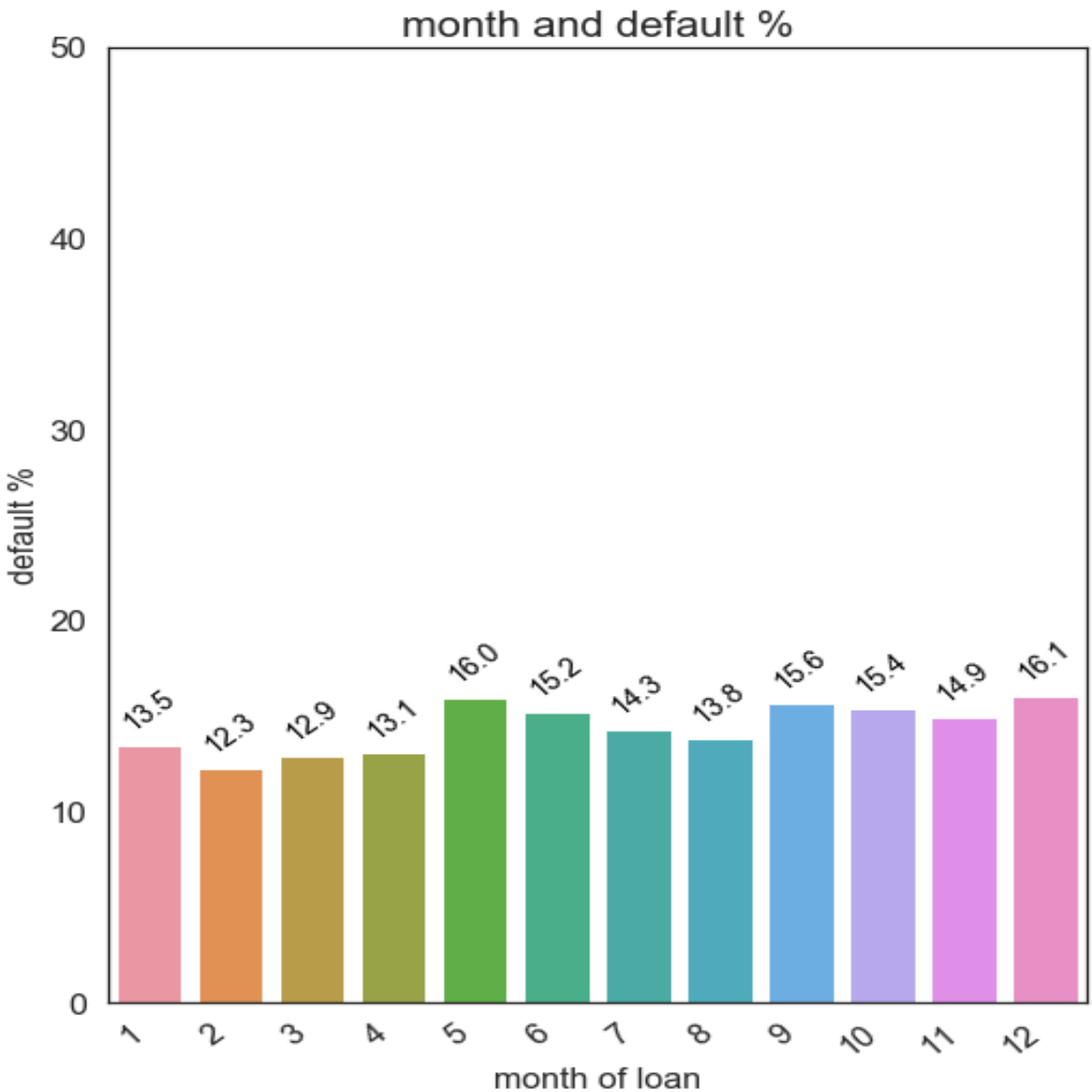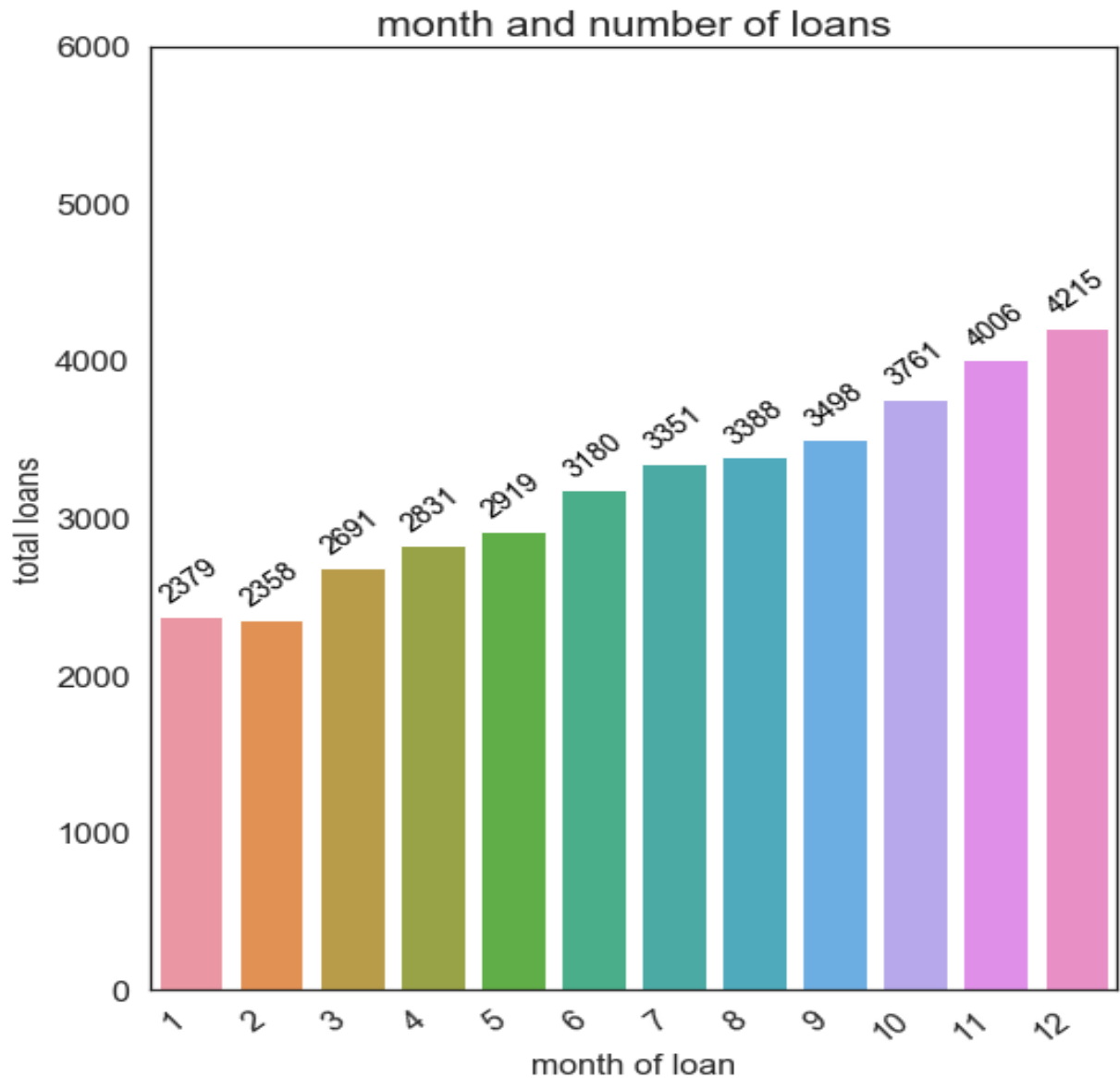
Default rate increases from 12.%2 to 16.7% as debt to income ratio increases. Business volume also high at higher DTI.
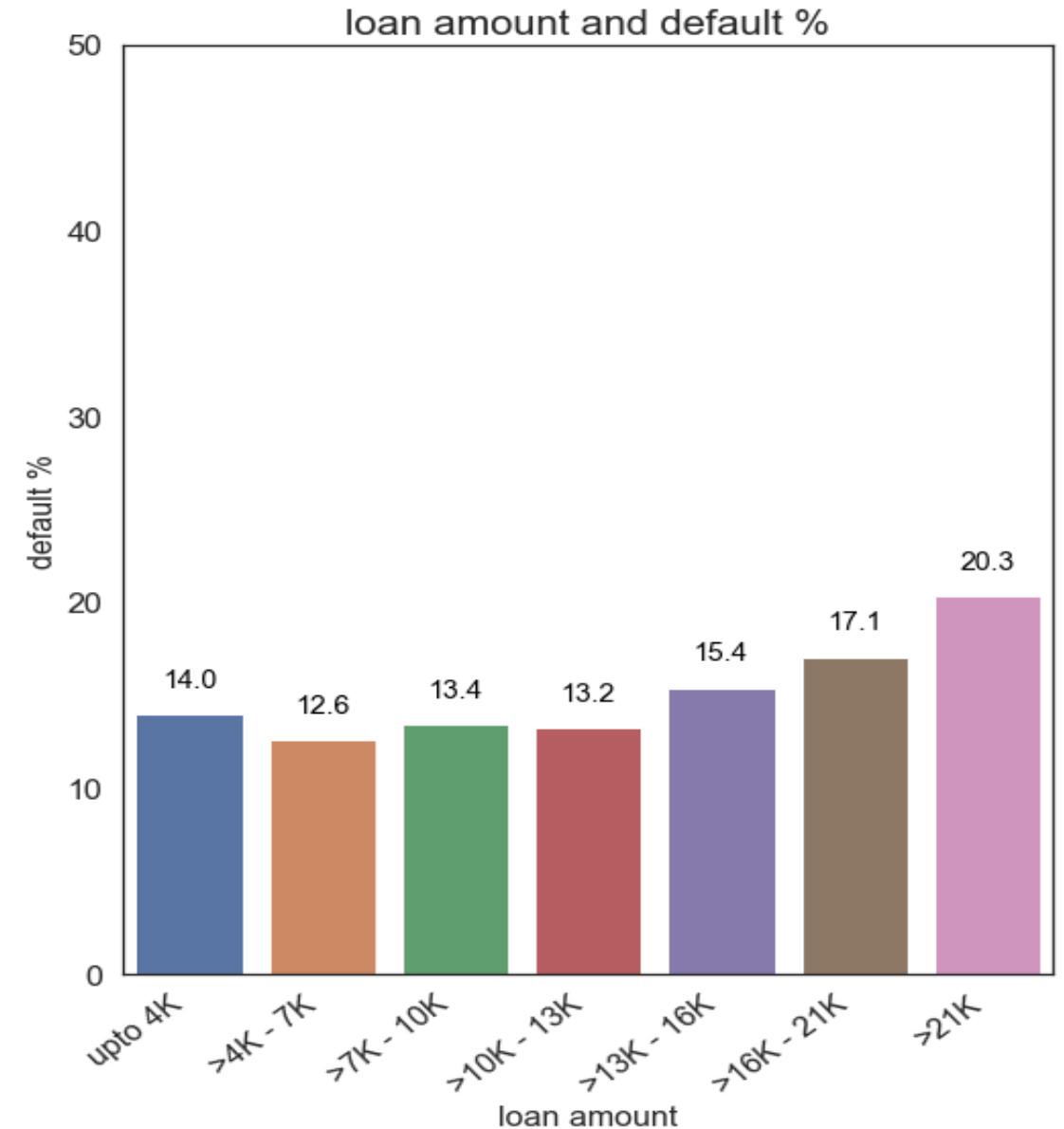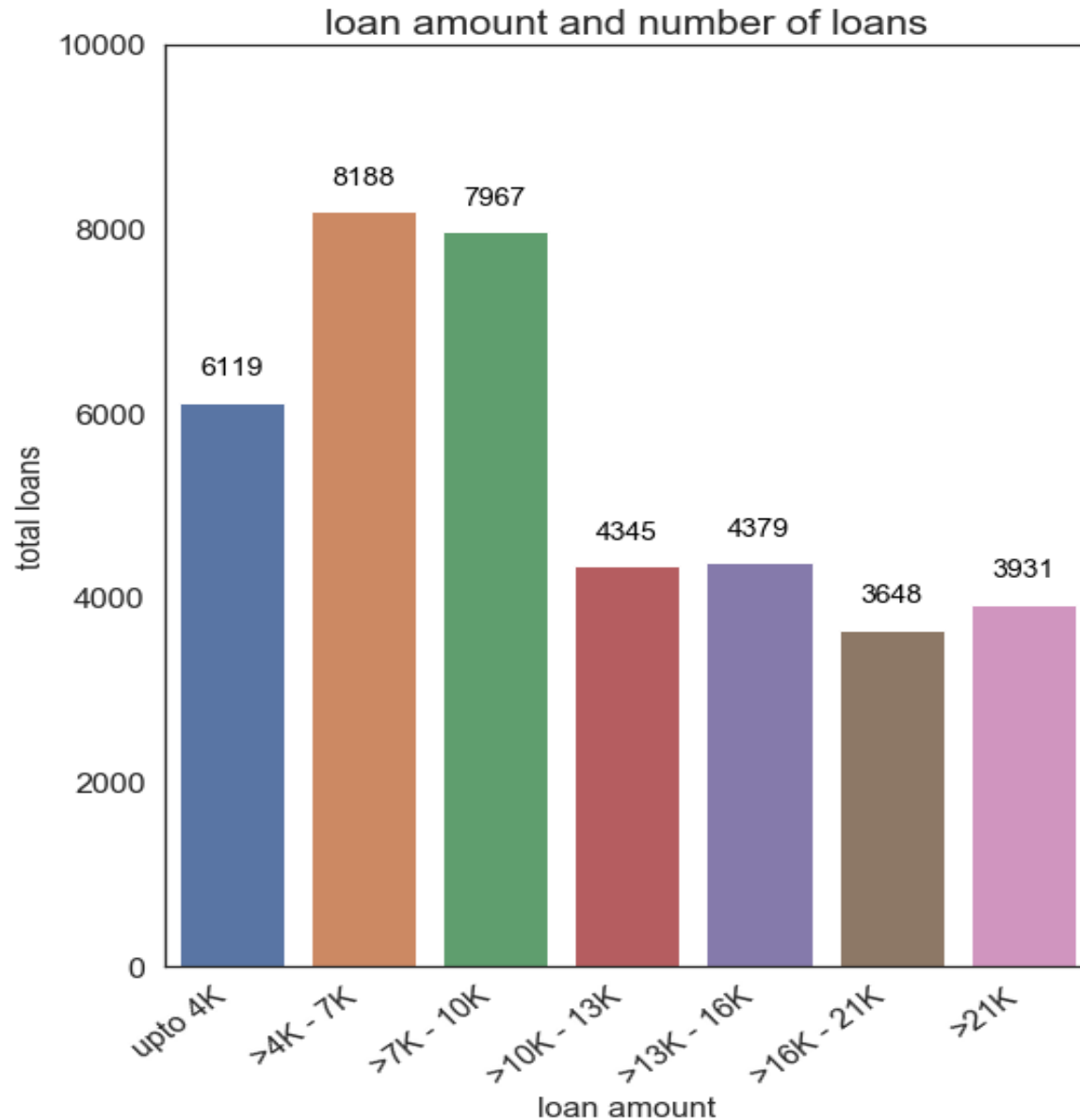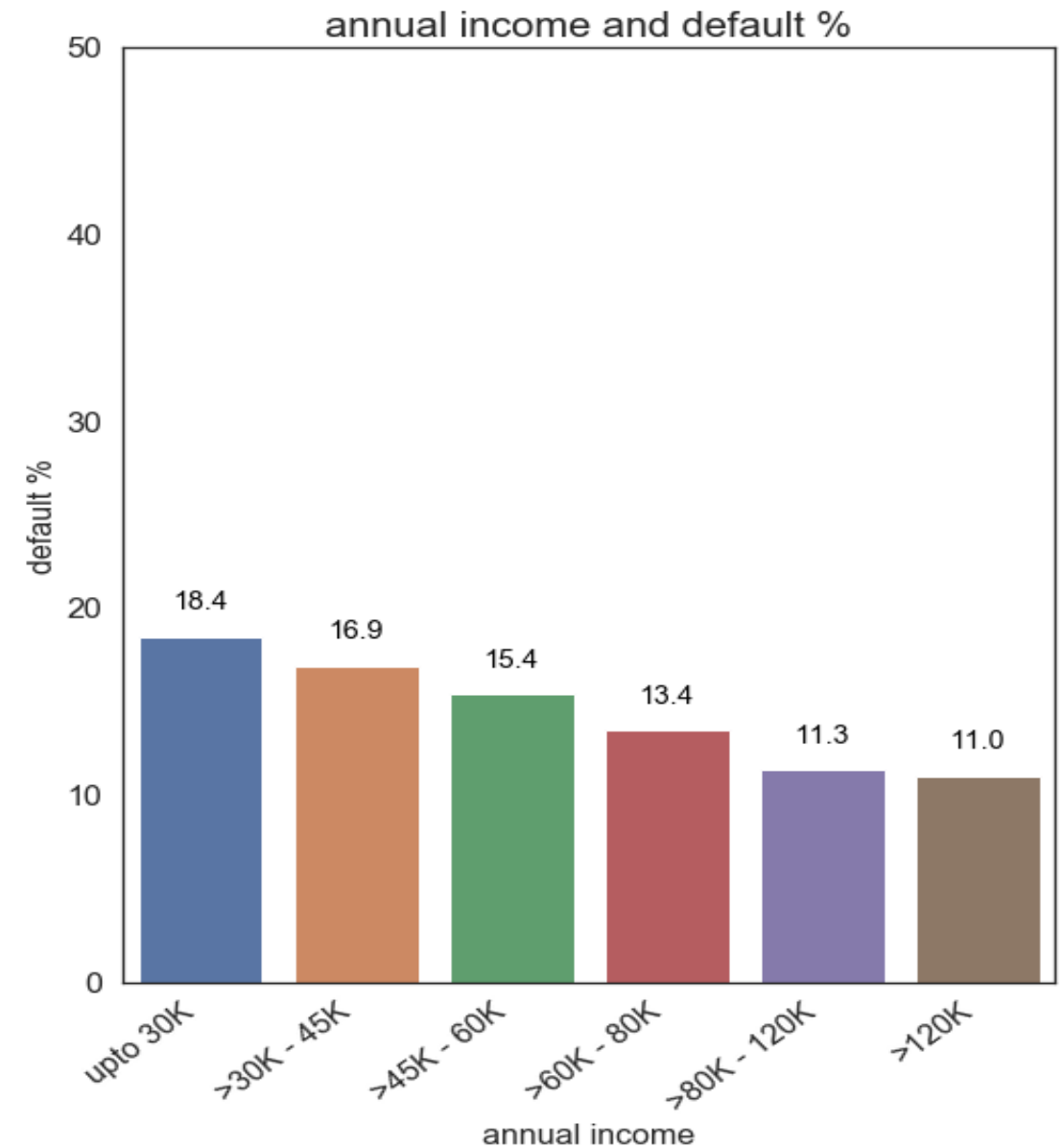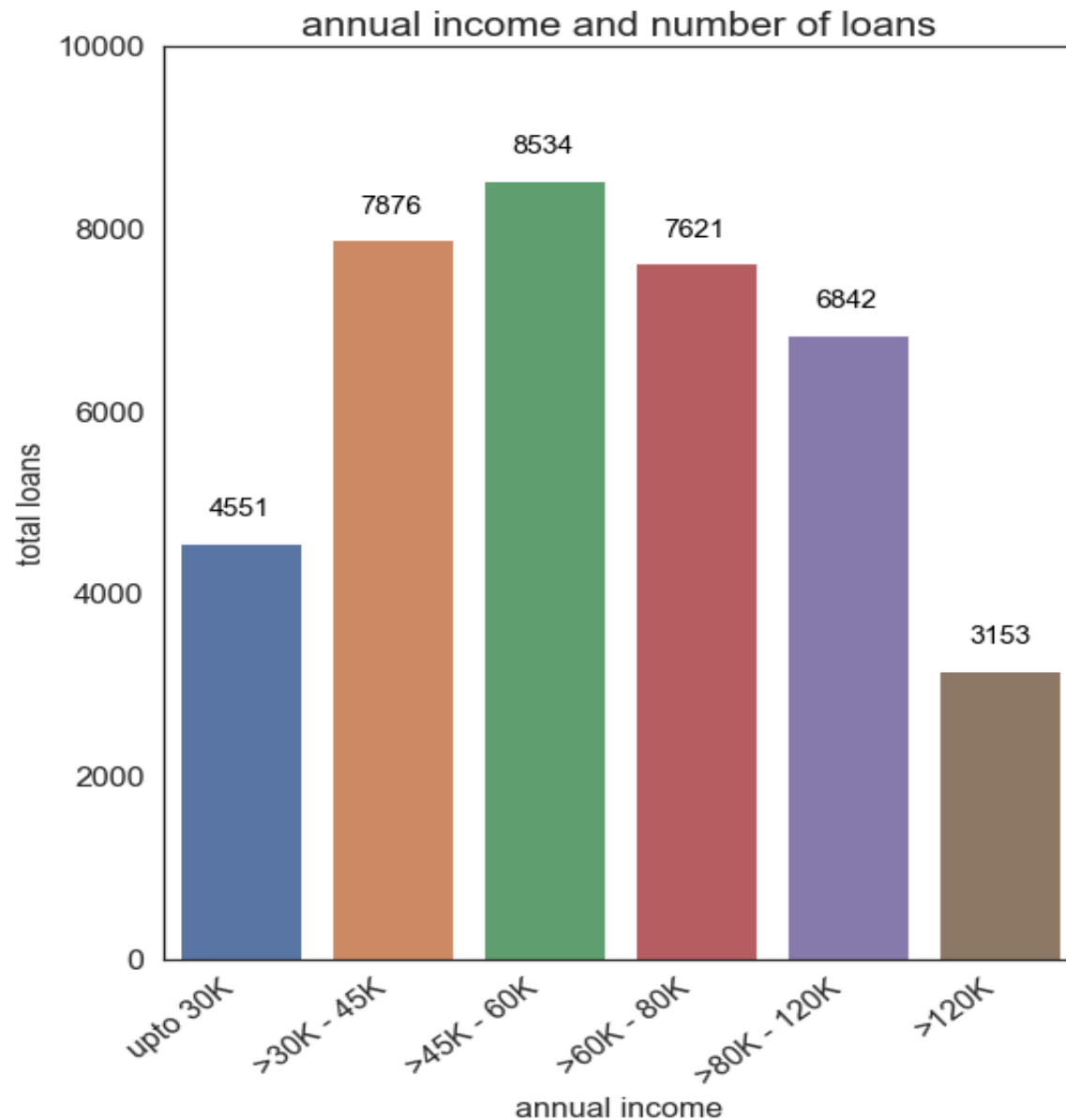
Default rate higher in certain months. December in particular has high business volume and high default rate.
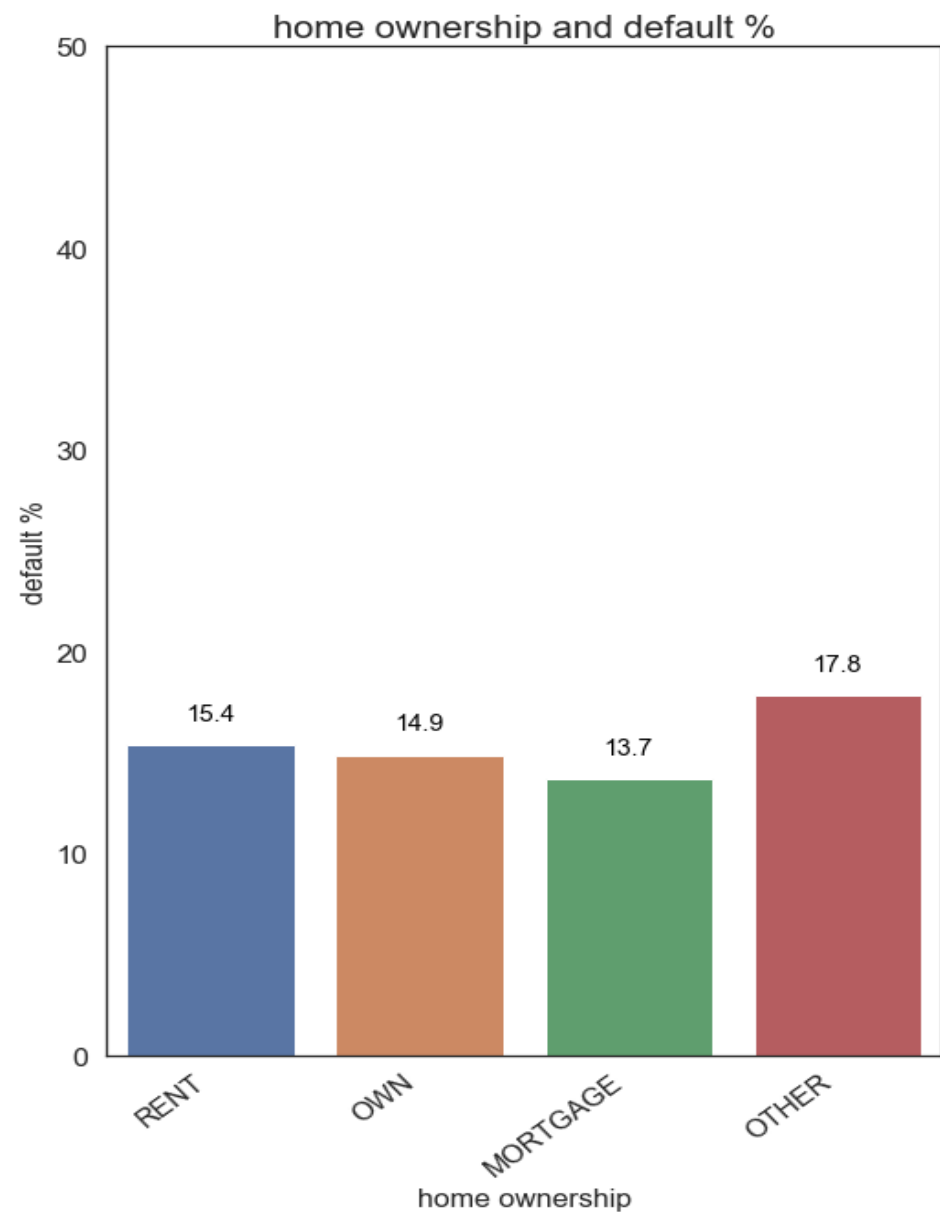
Higher the loan amount, higher the default rate.

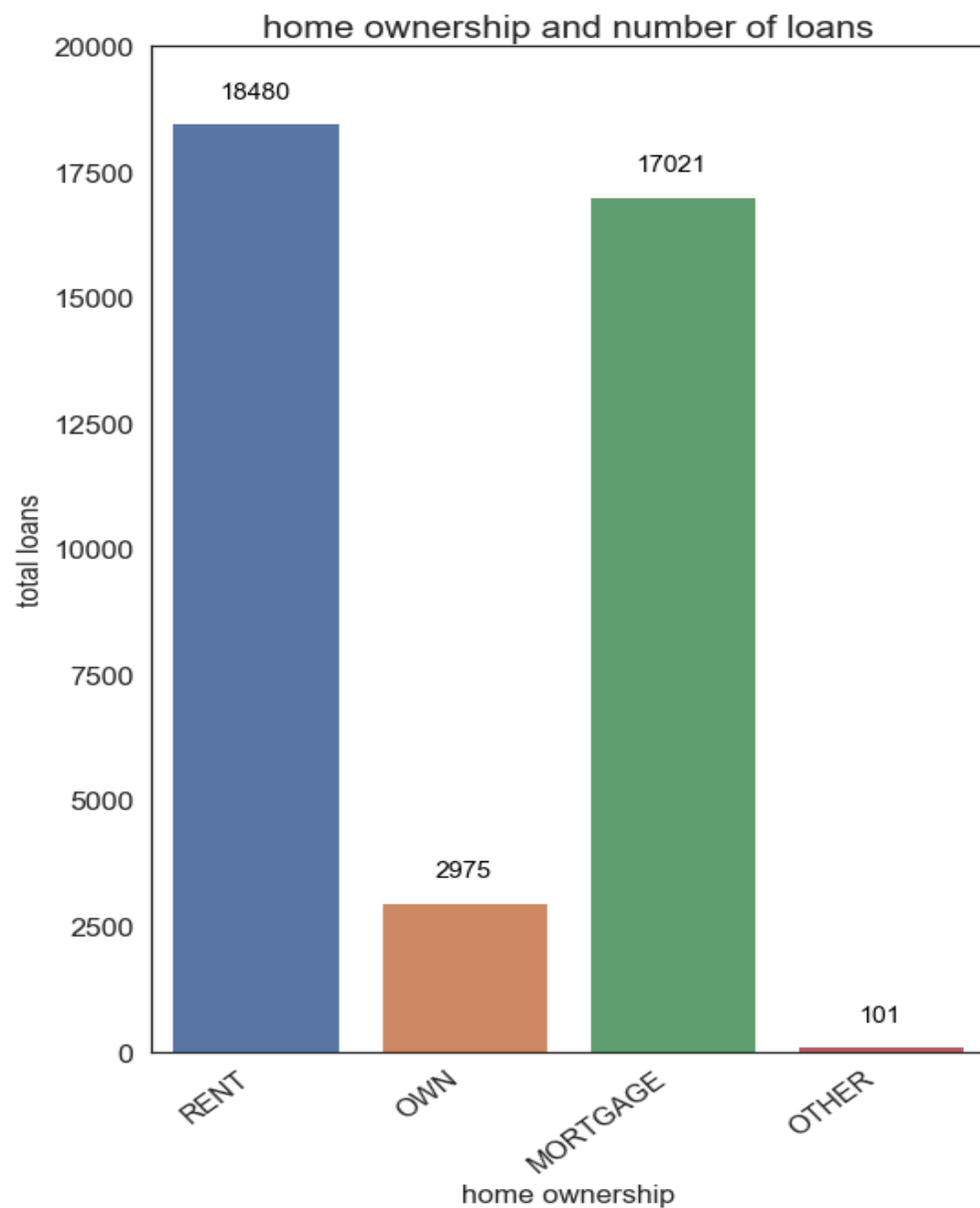Lower the income, higher the default rate.

No significant link between type of home ownership and default rate

home ownership and number of loans

home ownership	total loans
RENT	18480
OWN	2975
MORTGAGE	17021
OTHER	101

home ownership and default %

home ownership	default %
RENT	15.4
OWN	14.9
MORTGAGE	13.7
OTHER	17.8

No significant link between employment duration and default rate
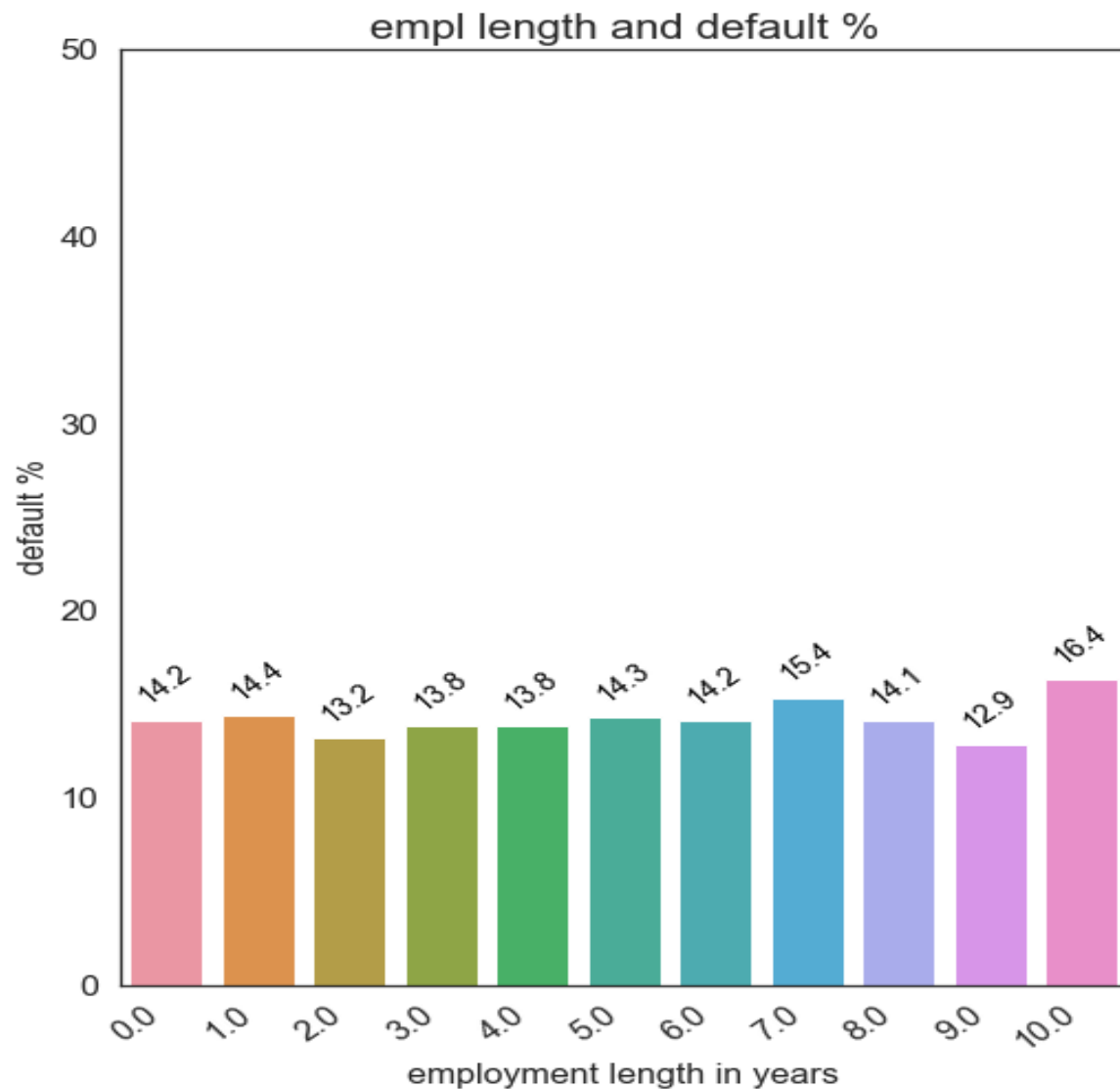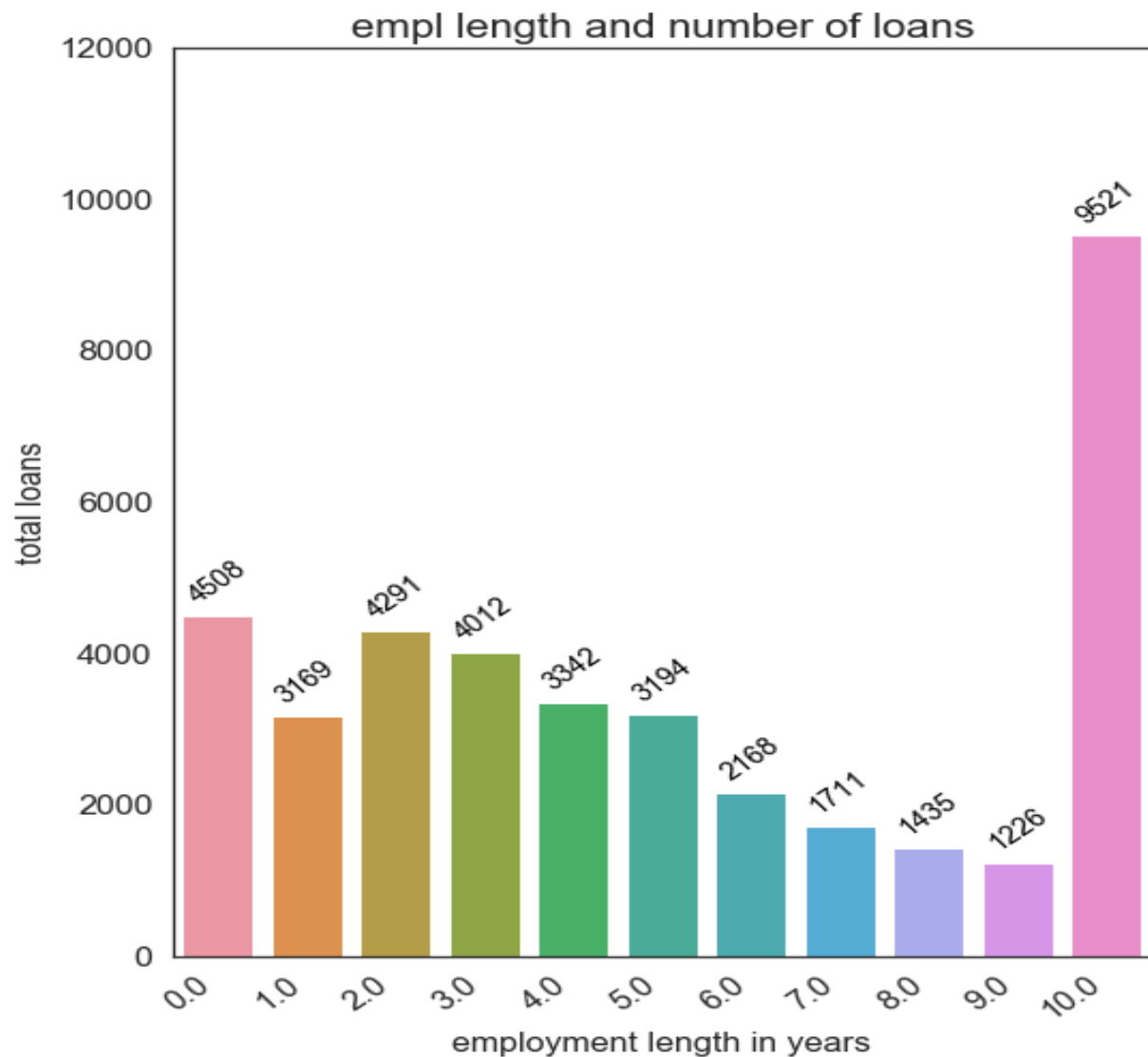
10.0 refers to employment service 10 years of longer

State and number of loans

State and default %

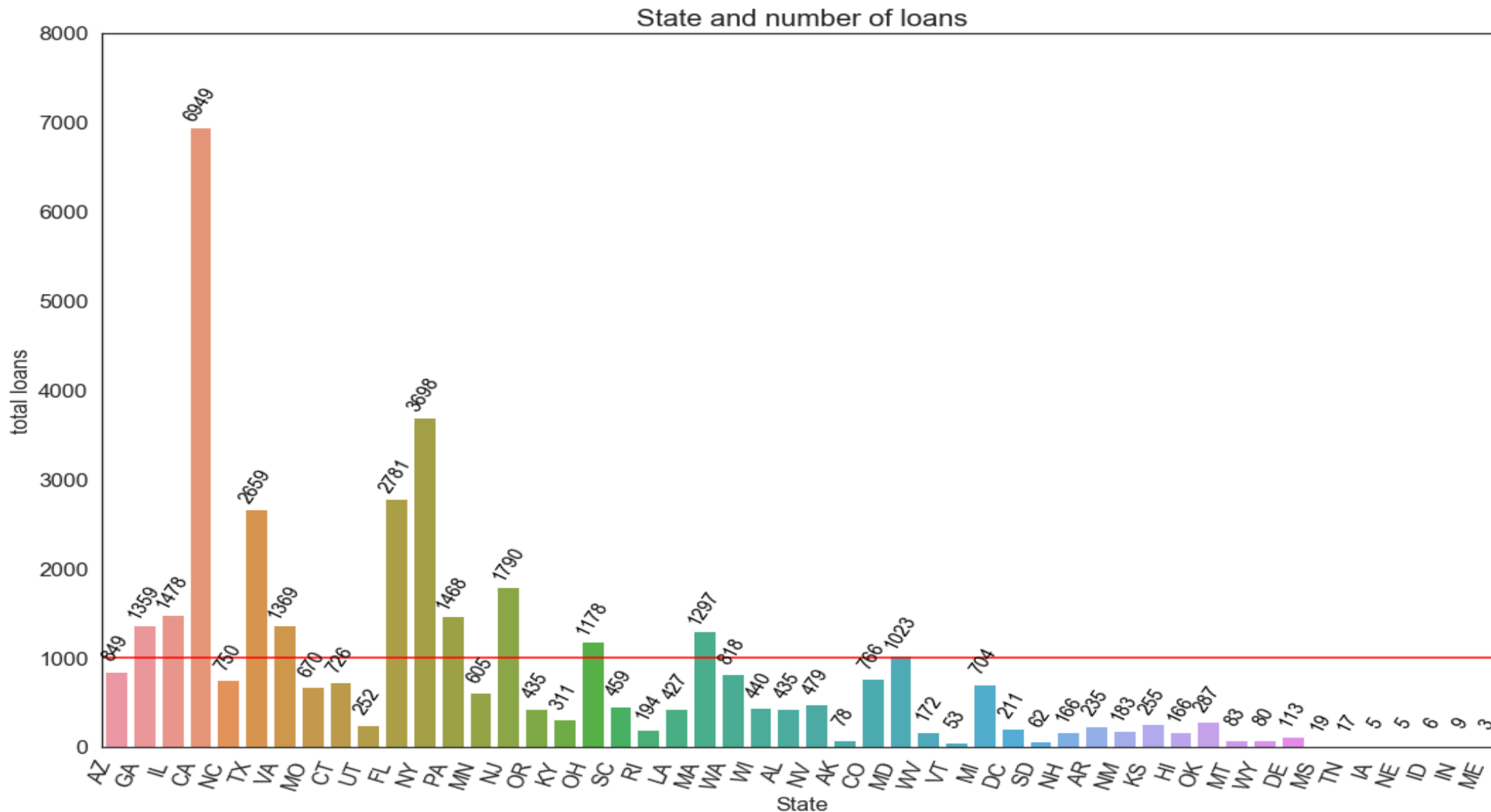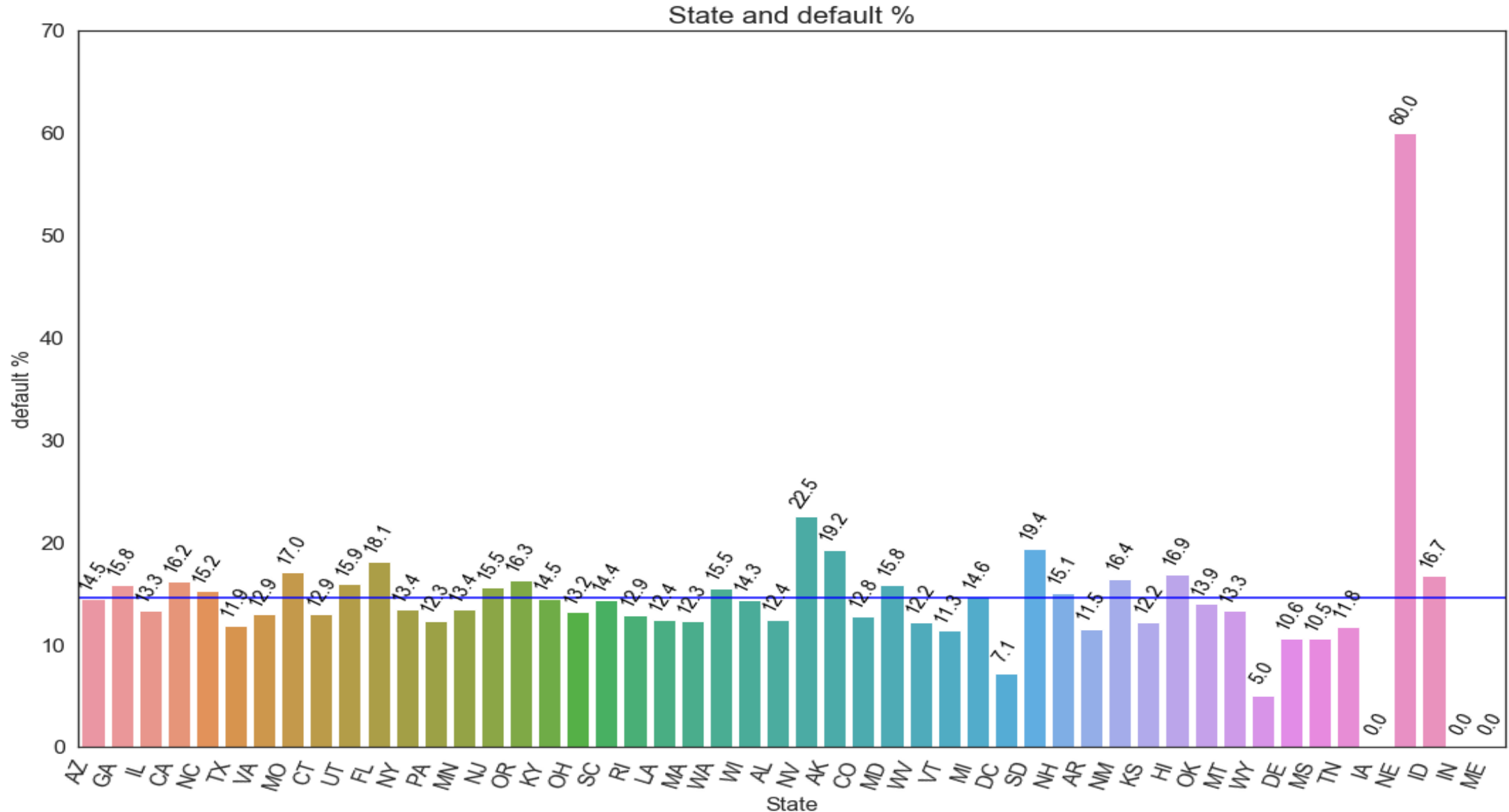47% of loans given are for "debt consolidation", not in top 5 default rates. **UpGrad**
Small business has very high default rate.
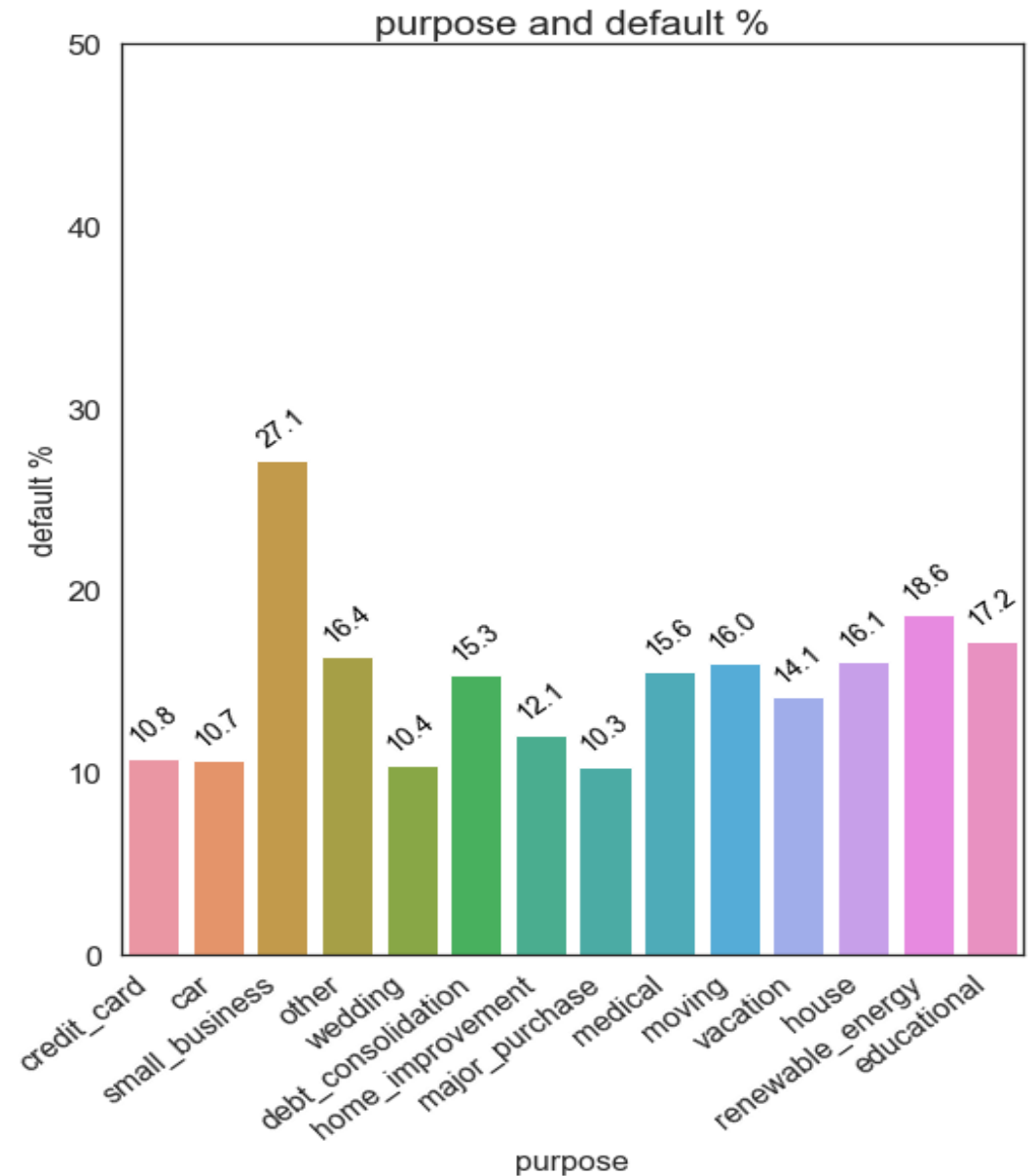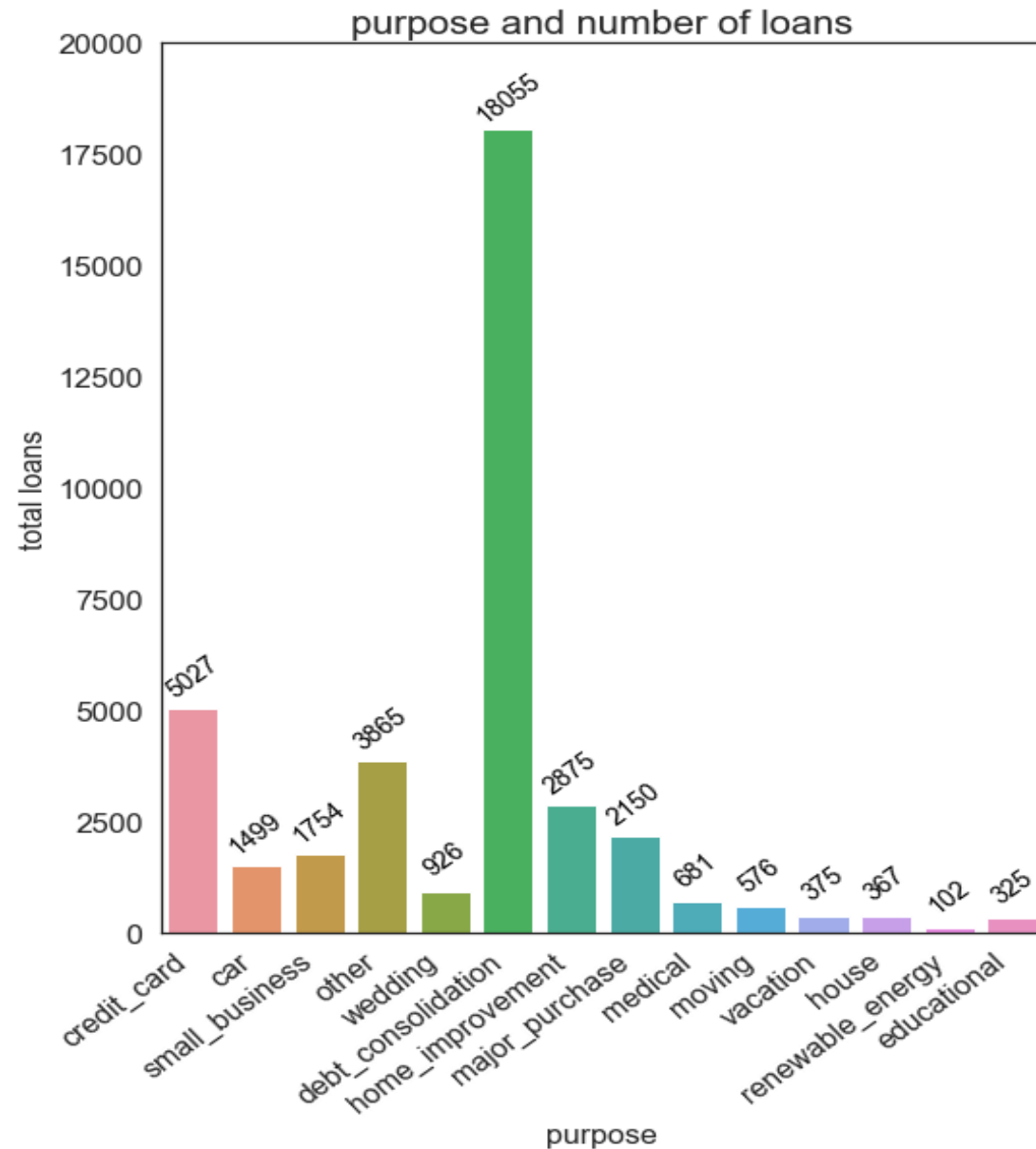
Bivariate Analysis

Amongst the shortlisted variables from univariate analysis, let us explore for any correlations between two variables for default rate

Variables identified for bivariate analysis

- Term
- Grade*
- Interest rate*
- Debt to income ratio
- Loan amount
- Month of loan application
- Annual income (upto 150K, to exclude outliers)
- Employment length
- State (CA, TX, FL, NY, NJ, IL, PA, GA)#
- Purpose (small business, other, debt consolidation, major purchase, home improvement, credit card)#

(*)these could variables derived by Lending Club basis many variables; these could therefore be expected to have high correlation with default rates; (#) these have sufficient volume and higher default rate

Analysis 1

No significant correlation observed between pairs of income, dti, loan amount, interest rate, employment length, month of loan and term. Correlation between income and loan amount by term and loan status did not reveal any significant change with correlation ranging between 0.4 and 0.47.

Heat map
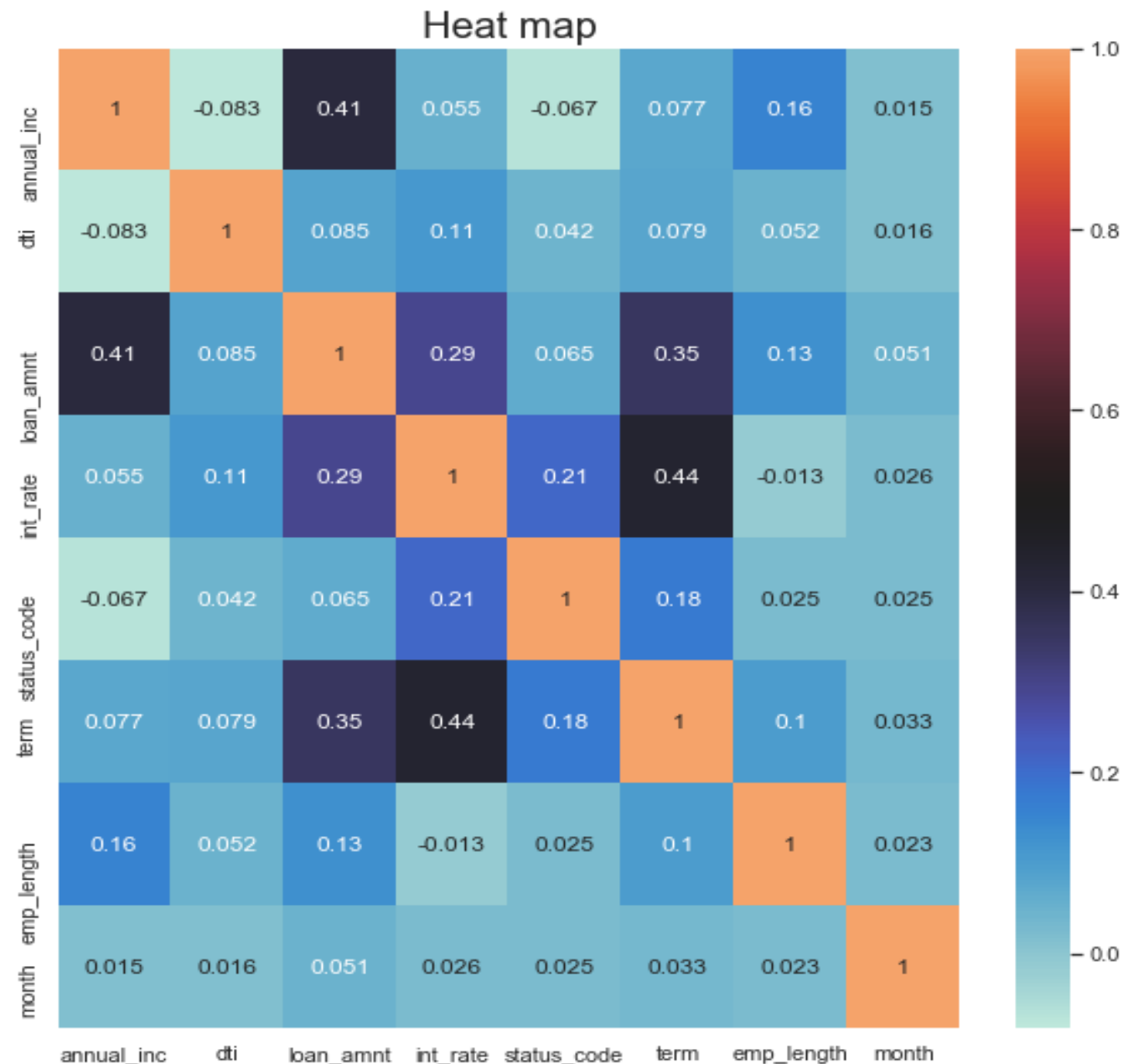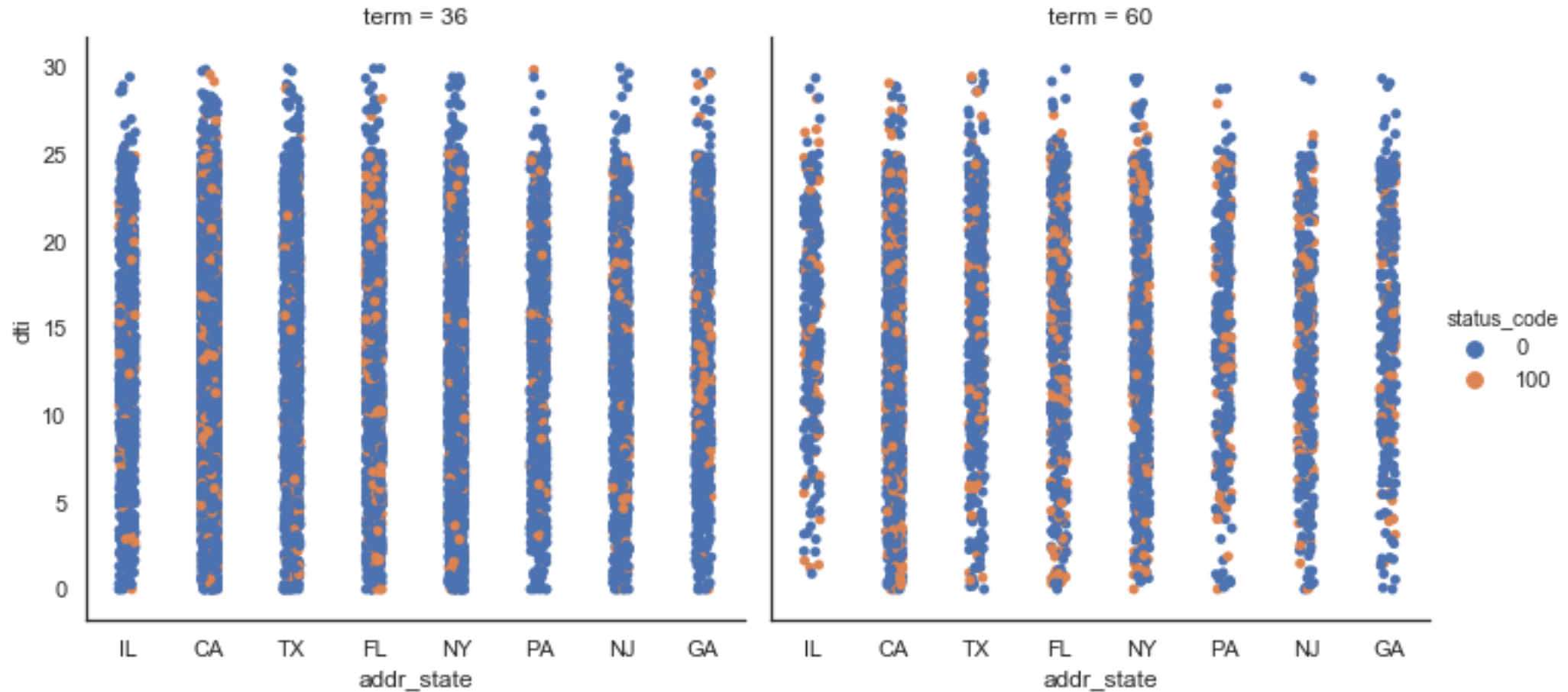
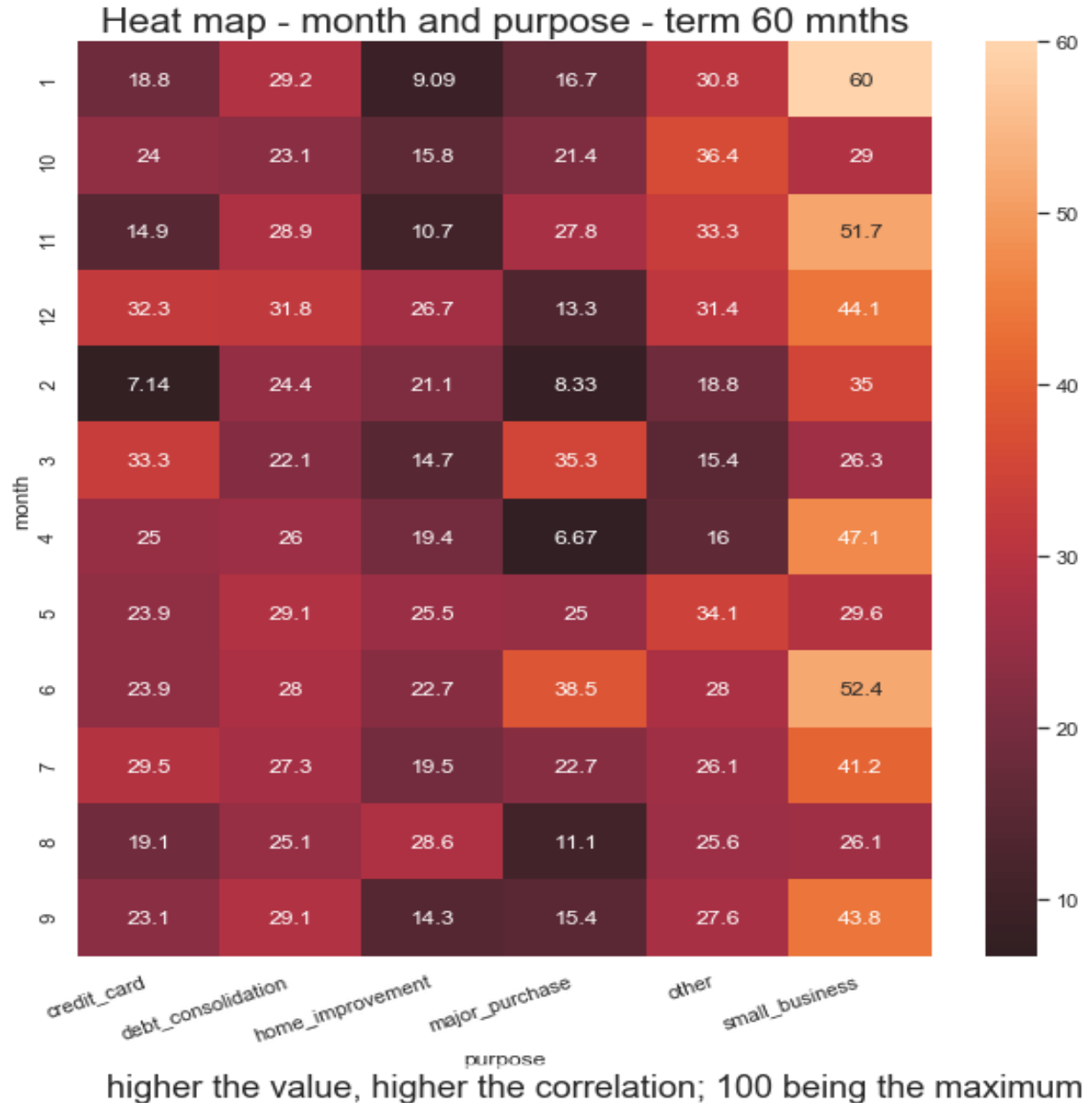Analysis of above kind between pairs of grade, address State and purpose with interest rate, dti, loan amount and income reveal higher default risk for term 60 months across these.

Status Code 100 means defaulted, 0 means fully paid

Analysis …. 3

No significant correlation observed between pairs of grade, address State, purpose, month of loan and employment length in the heat maps for these analysed for term 36 and term 60.

Heat map - month and purpose - term 60 mnths

higher the value, higher the correlation; 100 being the maximum

Conclusions

Applicants with one or more of these characteristics carry a high risk of default:

- Term 60 months

- Loans applied in months of May and December

- Debt to income ratio of above 19

- Loan amount above 16,000

- Income below 45,000

- Address in FL or CA (and a others with quite small business volumes)

- Purpose of loan is "small business" – high risk

- Higher grade and higher interest rate (these are perhaps high because the applicant is credit risky)