Lending Club Case Study

SUBMISSION

Name:

Gayathri **C**

CHANDRIKA **PATURI**

Overview:

Background:

- **Lending Club** 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**

Business Objective:

- Company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators to identify the risky loan applicants at the time of loan application

Strategy of Analysis:

- Understanding data

- Cleaning the data

- Analysing the data

- Plotting the data

- Concluding the analysis
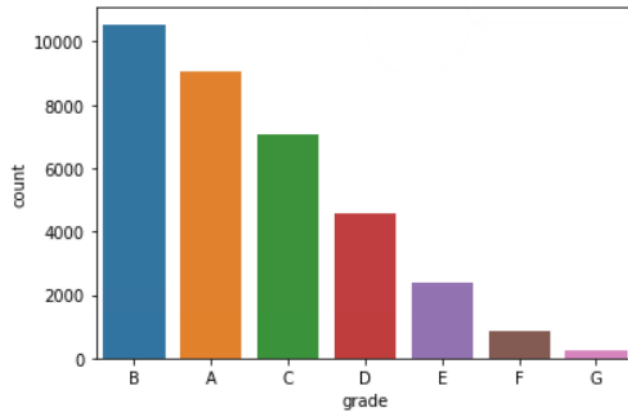
Problem solving methodology

Flow Chart
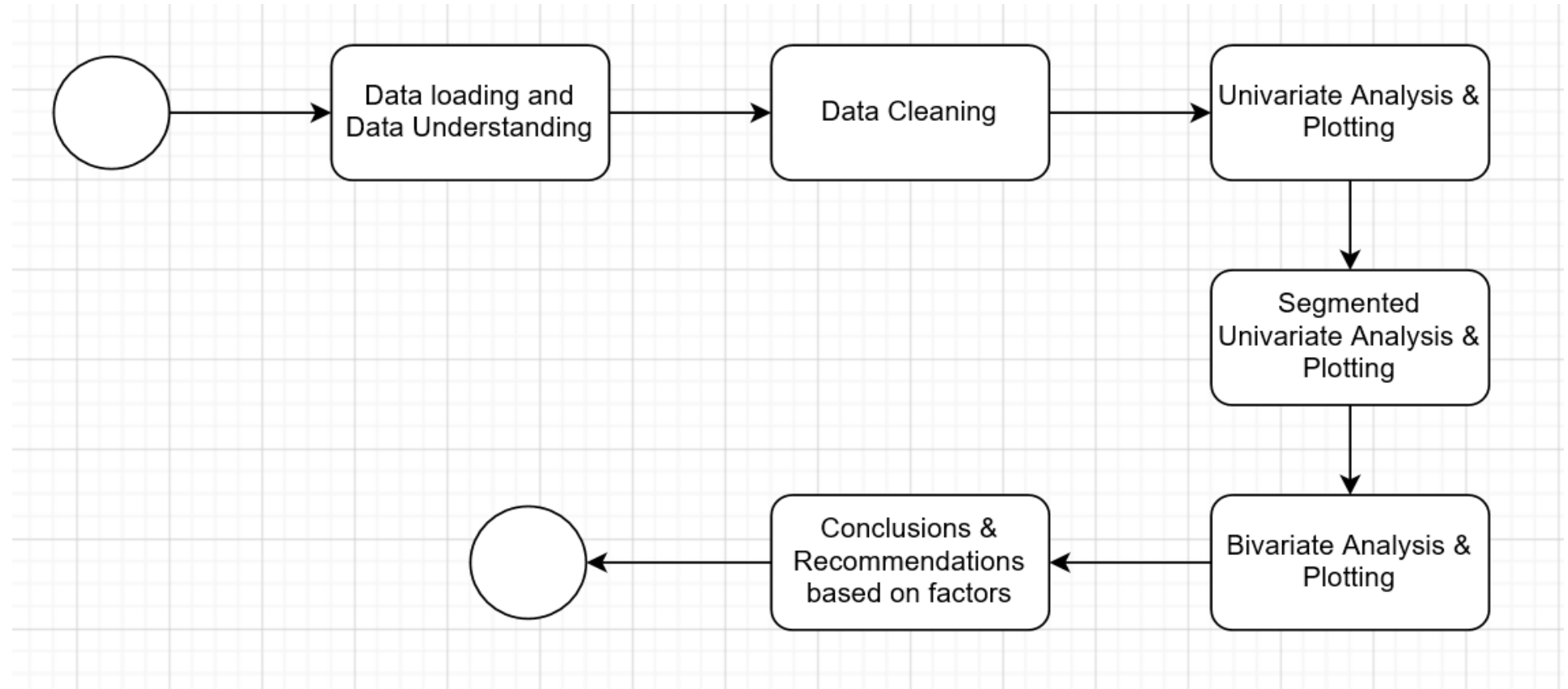
Analysis

- Loading the csv

- Understand the data

- Data cleaning – removing missing entries, converting the datatypes, data extraction and conversion, finding out the outliers and ignoring them, removing duplicate columns / rows, removing single valued columns etc
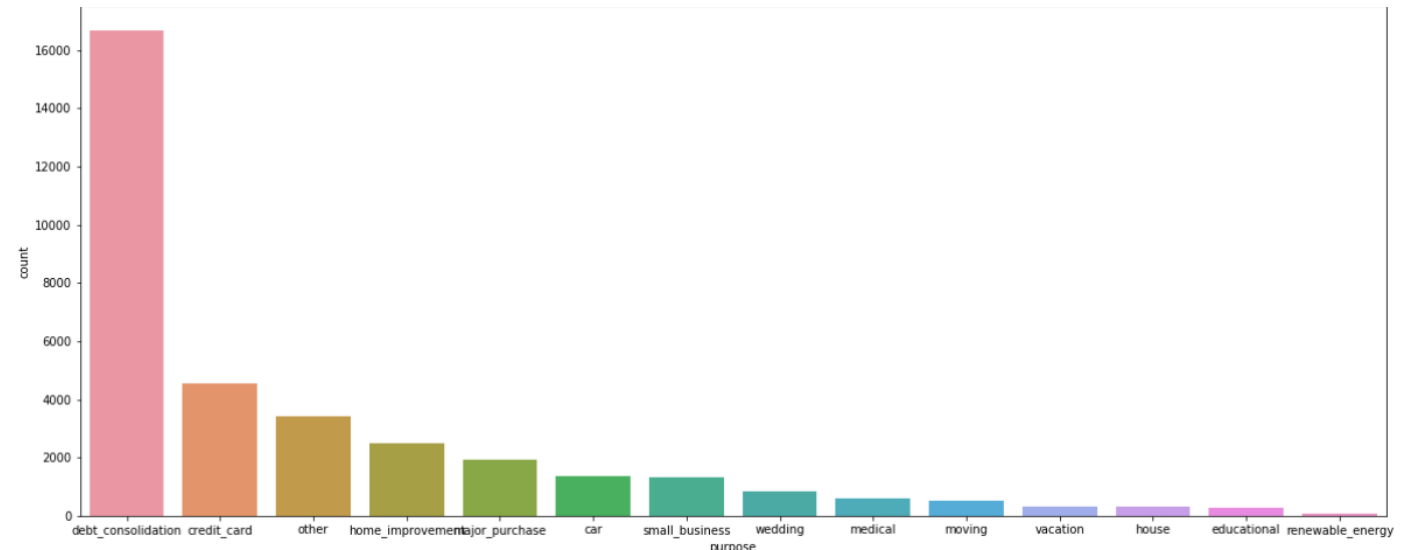
Univariate Analysis

Number of loans against grade

Higher number of loans are taken
by higher grade borrowers

Number of loans against purpose

Higher number of loans are taken for debt consolidation. Based on it, a bank or finance lender
can decide more offers to attract customers and give competition to competitors.

Number of loans based against emp length

Number of loans based on issue year
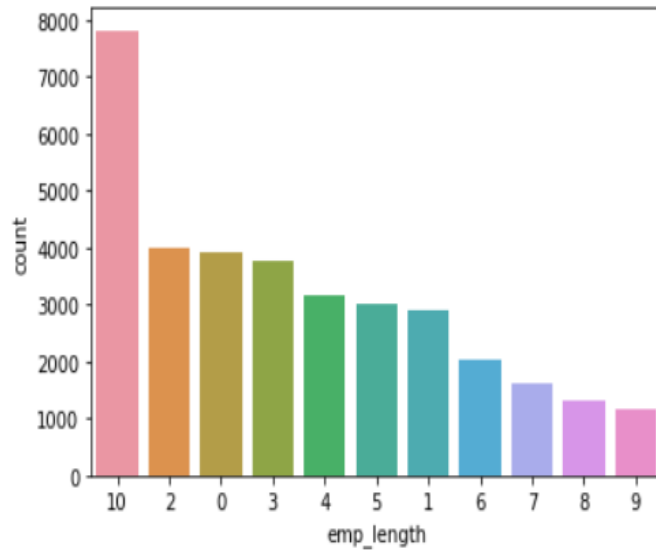
Higher number of loans are taken by borrowers with emp length 10 or above.
With #years of experience, the salary will get increased and the person can afford to take loan compared with a fresher or medium experienced person.
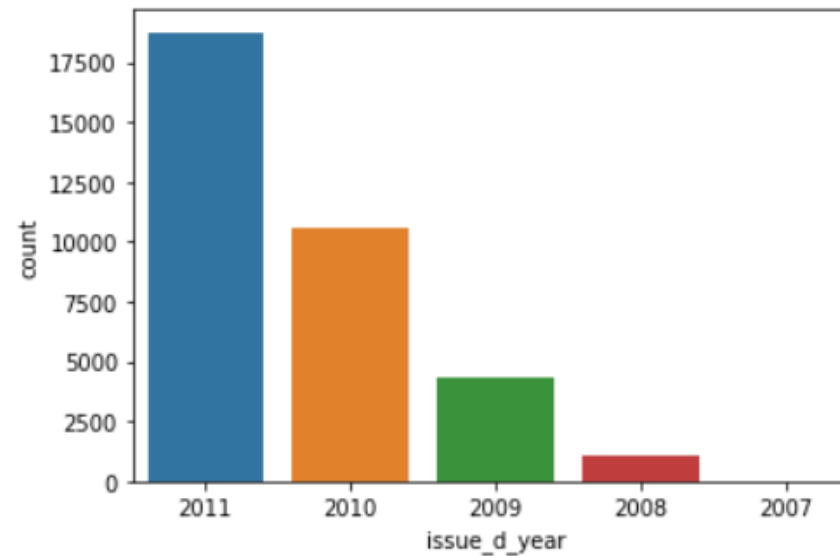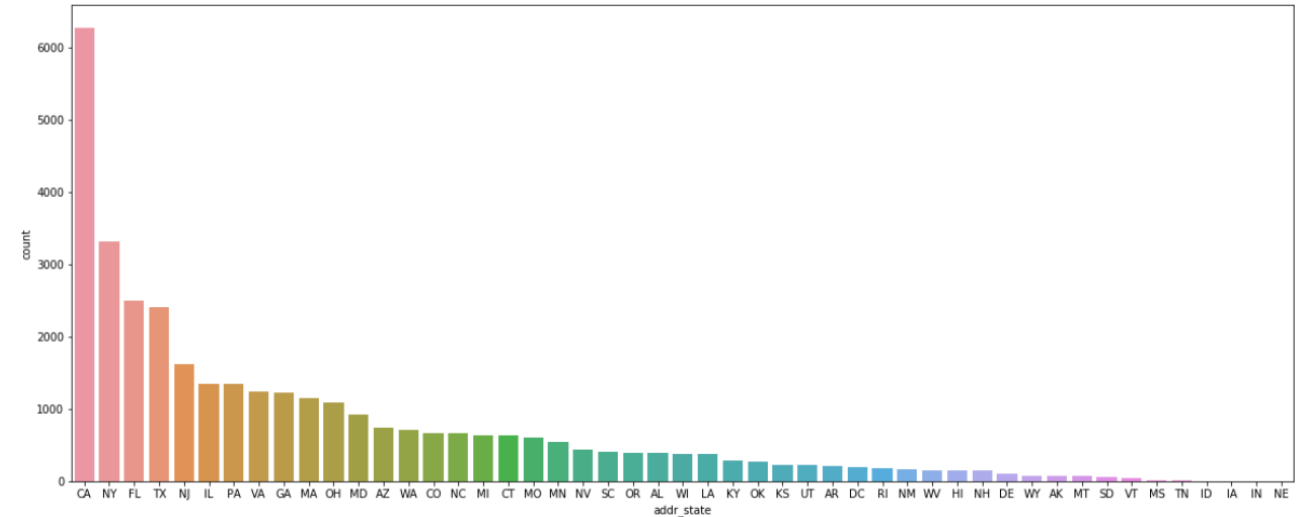
Higher number of loans are taken in 2011. Which determines the bank is running in profit with more #customers year on year.

Number of loans based on issue month
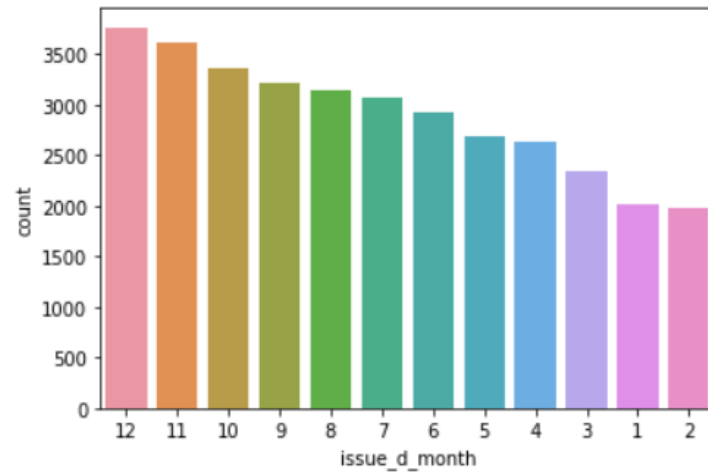
Higher number of loans are taken in Dec.
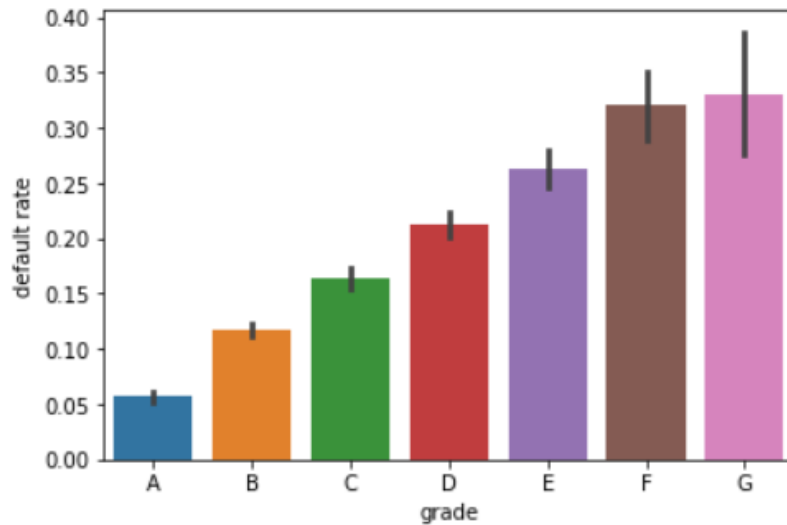Due to higher offers during festival time.

Number of loans by each state

Higher number of loans are taken by borrowers in state CA.
by looking at the above, lender can decide what to be done
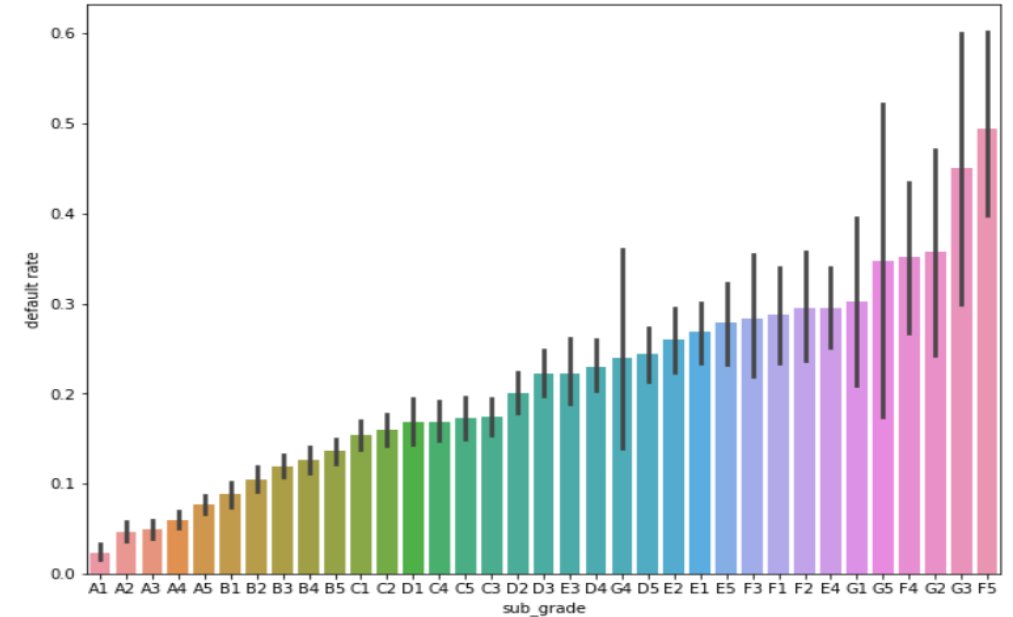so that the performance will be go high in other states too.

Segmented Univariate Analysis
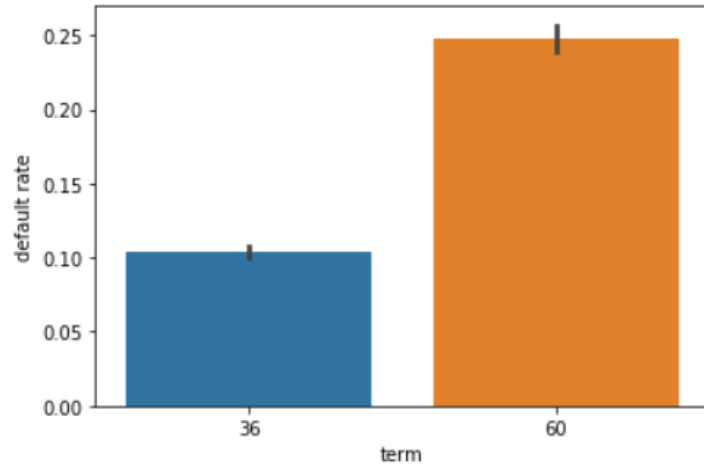
default rates across all grades.

default rates across subgrades

As the grade of loan (risky of loan) increases the default rate increases
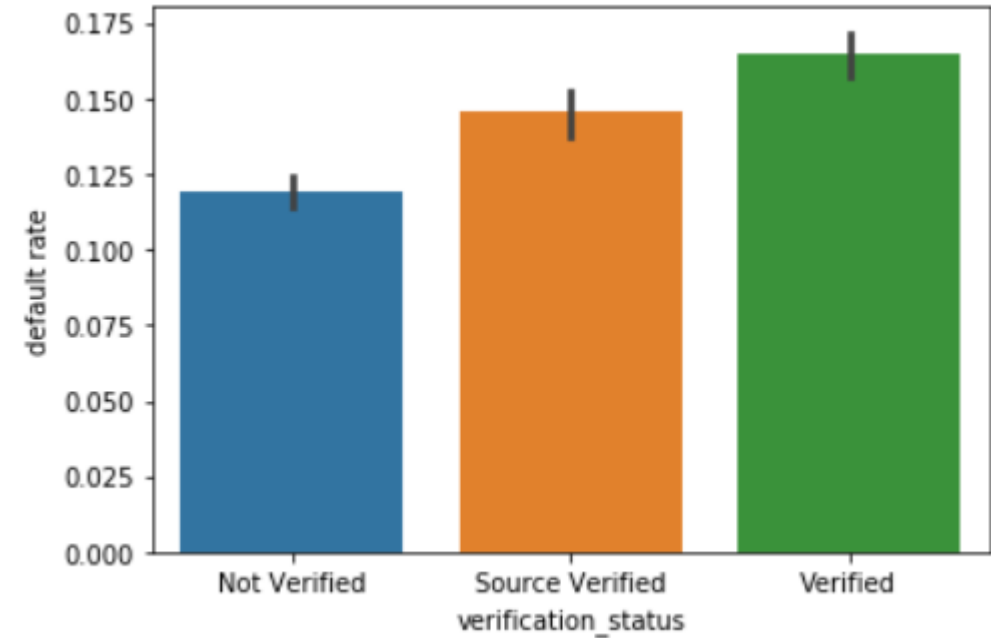
In-depth analysis which gives idea which sub grade is riskier than other grade.
For example: A1 is less riskier than A2. A2 is less riskier than A3 and so on...

default rates across monthly terms

The risk is more, i.e, defaulters are more when the term to repay is 60.
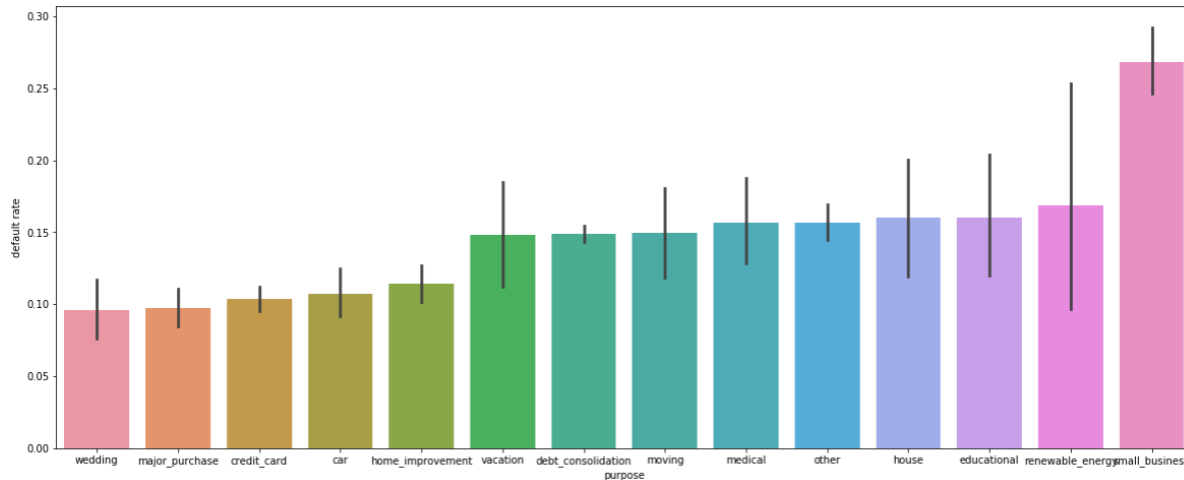

default rates across verified loans

verified users have more defaulters.
That means the verification process need to be corrected. Fraud during approvals of loans (could be bribing to get loan, could be relative / close knit to approve loan / submitting the fake proofs etc)
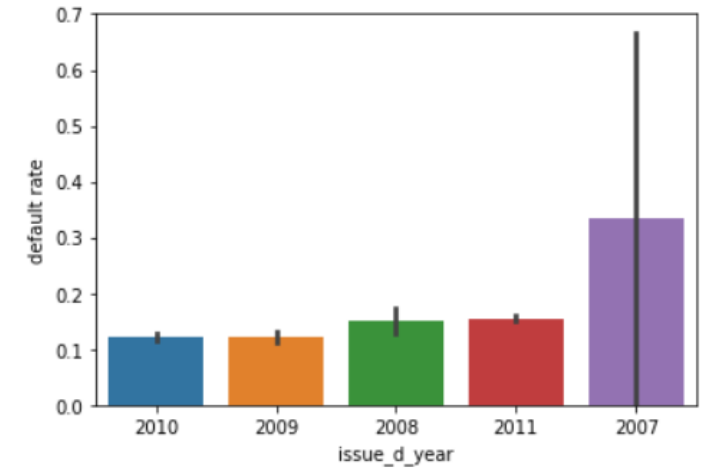
default rates based on purpose of taking loan

small business loans have more defaulters. Because,
1. in the world most of the people are middle class people
2. so, small business will be run by them
3. if no profitable business, may not have the financial condition to repay
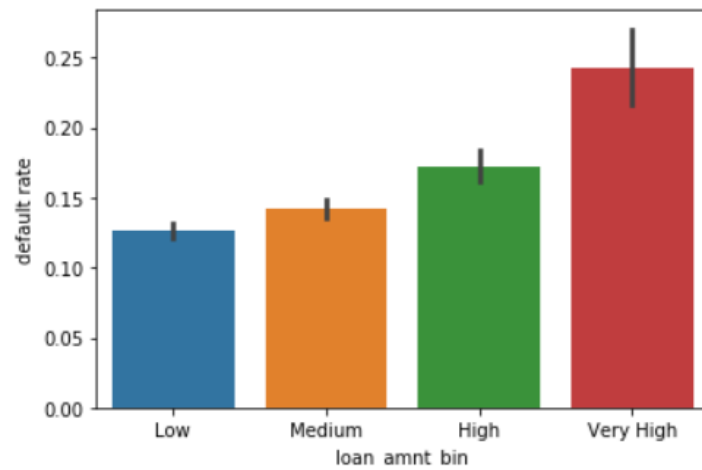
default rates across years & Months

In 2007 the defaulter's rate is more -
the lending club has been established in 2006. So, it is new experience and difficult to manage how to recover the loans from customers by lender.
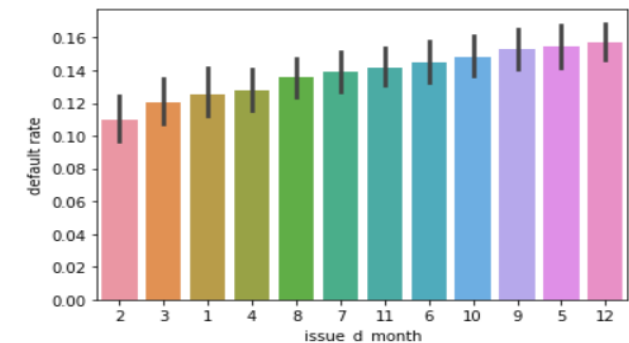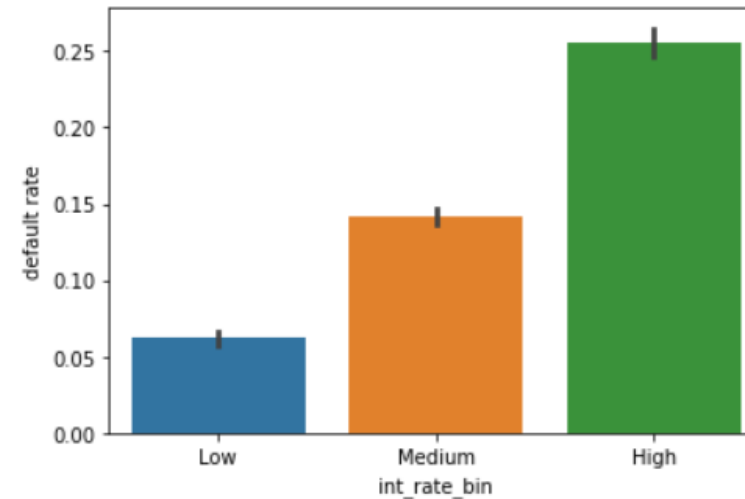

In December month the defaulter are more
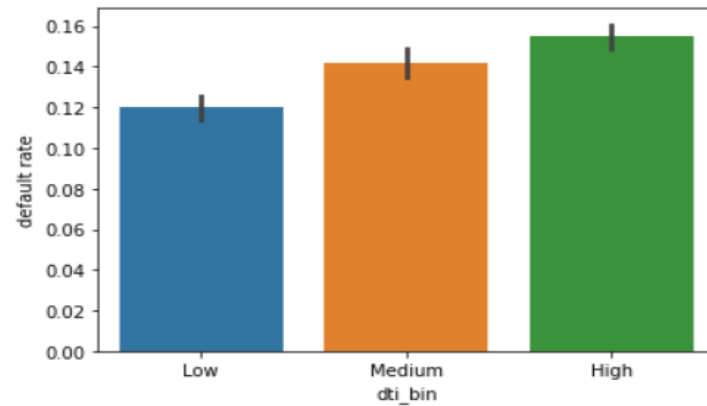
default rate across loan amount.

default rate across interest rate.

greater the loan amount, greater the risk with defaulters. because they may pay or may not pay huge amount.
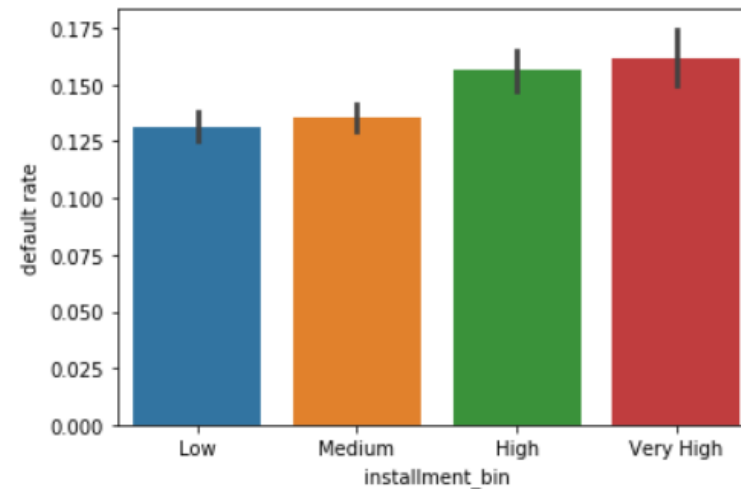
as the interest rate increases, the chance of having defaulters in the list.
As they may or may not pay the interest (if no profit in their business)
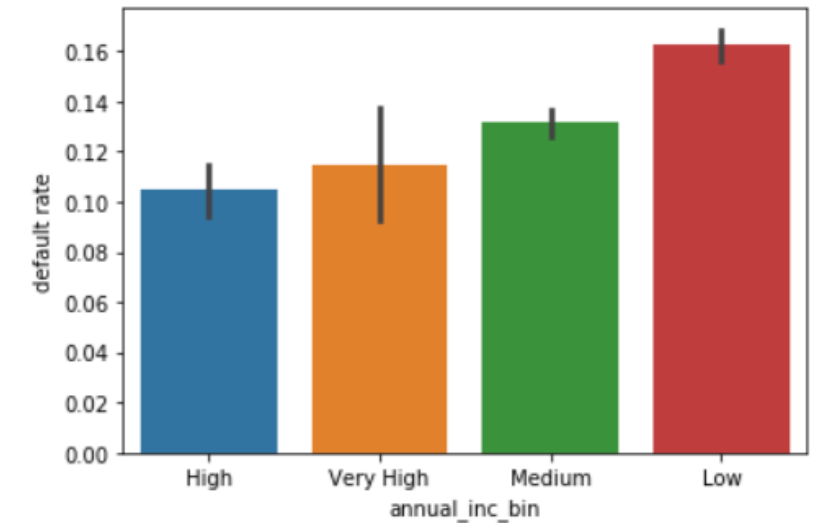
default rate across debt to income ratio.

higher the DTI, more defaulters.

default rate across instalments

higher instalments , then more defaulters.
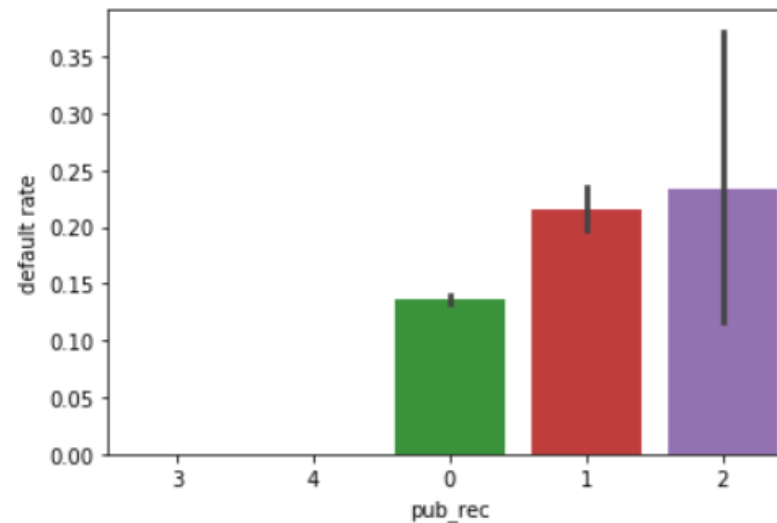
default rate across income

repaying the loan depends on annual income of a person / organization. So, lower the annual income, greater the risk by having more defaulters.
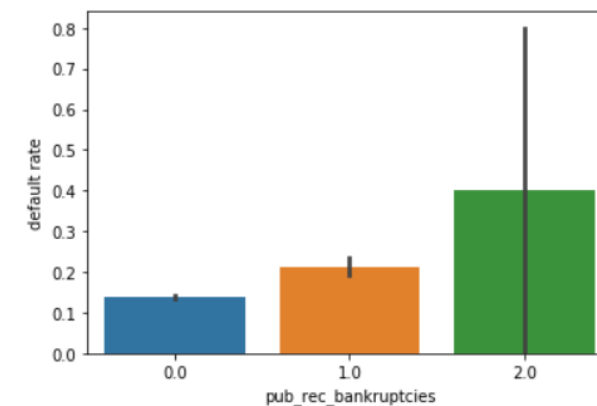
default rate across revolving utilization

default rate across public records

default rate across public records
bankruptcies
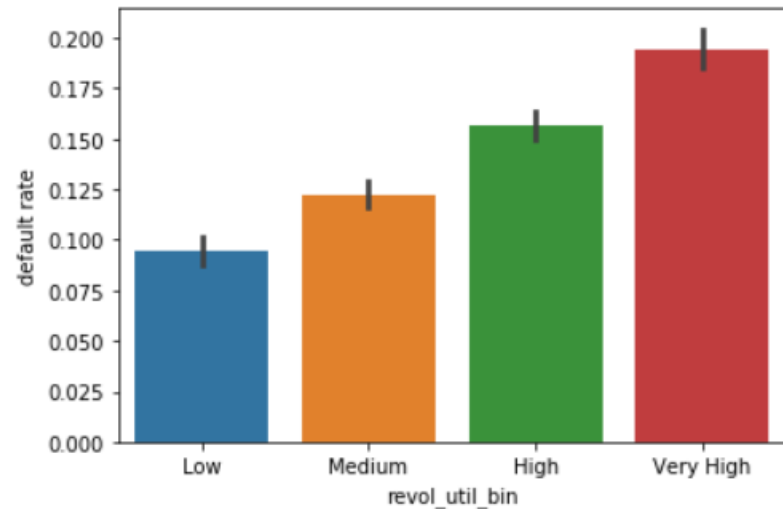
higher the revolving utilization, more risk to
receive the amounts from customers.

default rate increases with increase in
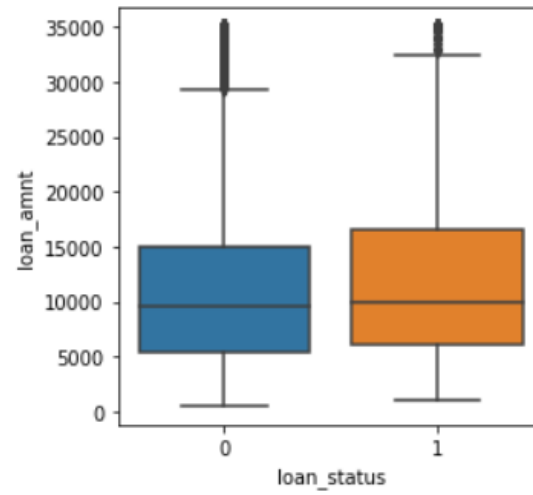number of derogatory public records

default rate increases with increase in
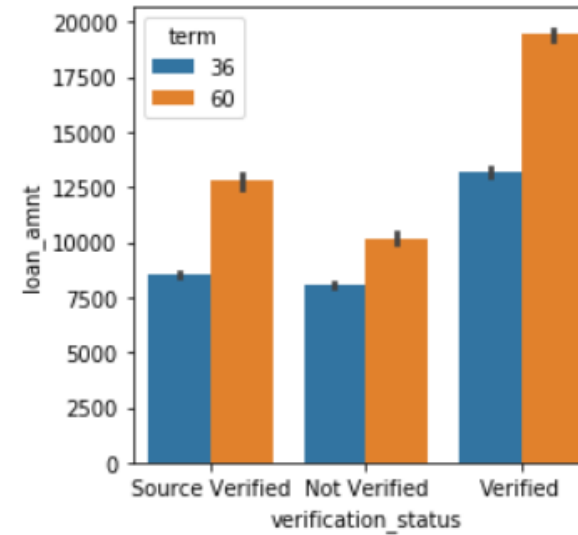number of pub_rec_bankruptcies

Bivariate Analysis
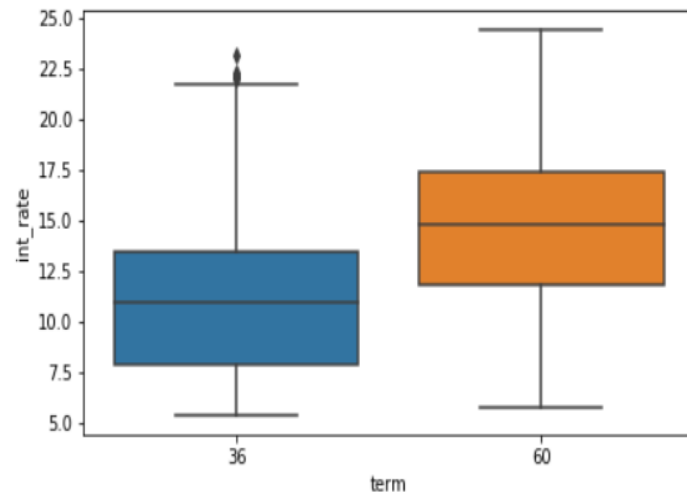
loan amount for each loan status

higher average loan amount gives more chances of
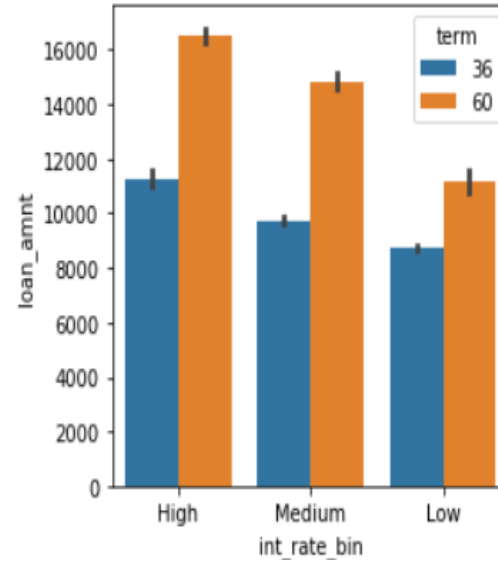defaulters.

loan amount against verification status

Higher loan amounts are more verified
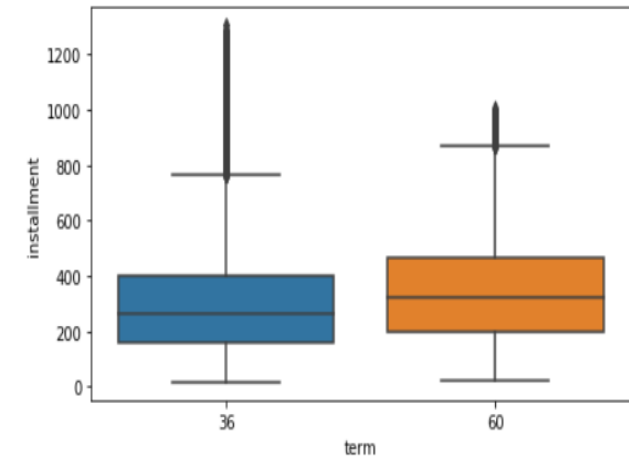
Interest rate against term

loan amount against interest

instalment against term
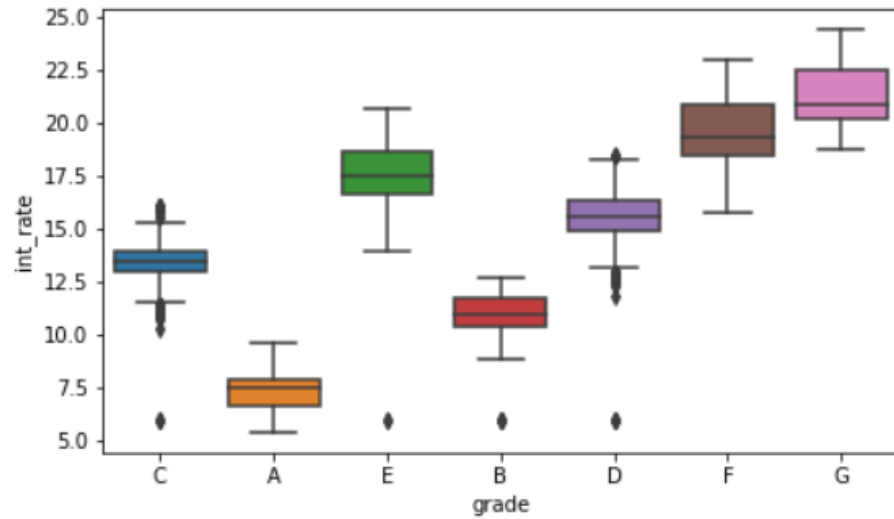
60 months term has higher interest rate compared to 36 months

Higher loan amount has higher interest rate depending on the tenure. by considering earlier factors, more tenure has more defaulters

60 months term has higher average instalment compared to 36 months
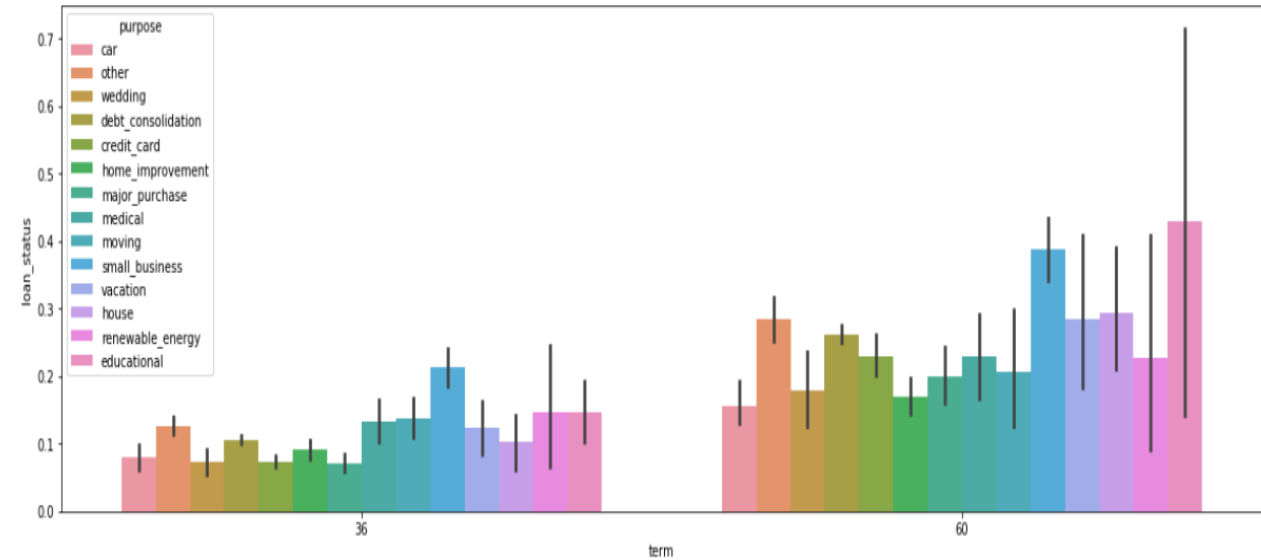
Interest rate against grade

default rate against term each purpose

lower grades have higher interest rate

Small business purpose loan are higher in 36 months term whereas renewable energy is higher for 60 months
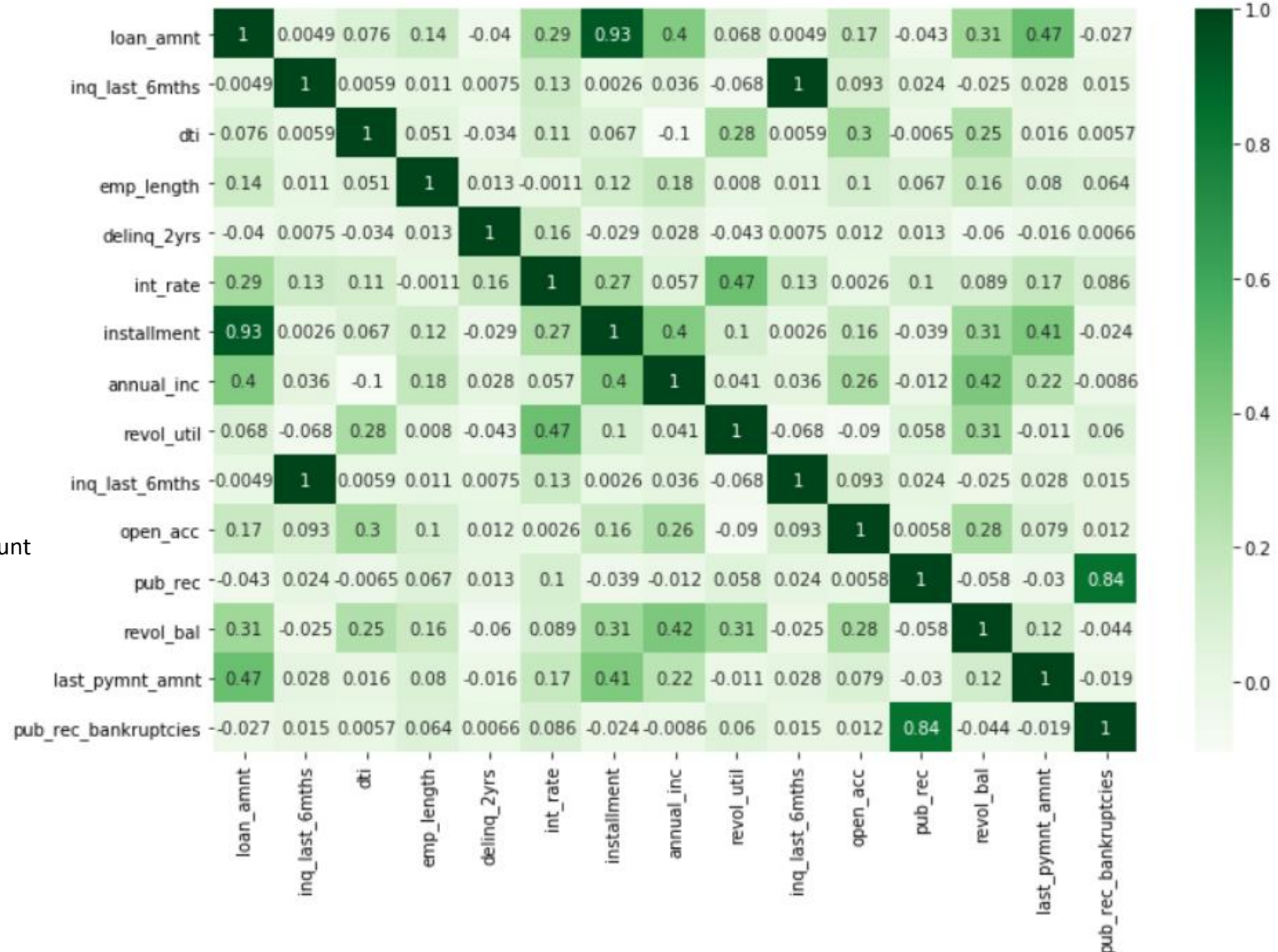
Heat Map to Find the Correlation between different numeric factors

Few of the findings are,

interest rate and revolution-util is correlated by a factor of 0.47

loan amount and last payment amount is also correlated by 0.47

etc....

Conclusions

After our analysis, below are the Major Factors to identify the risky loan applicants at the time of loan application.

- term

- grade and subgrade

- purpose

- verification_status

- loan amount

- interest rate

- dti

- installments

- revolving util

- annual income

- public records

Recommendation

- Improve the verification process

  Reason: we are seeing more defaulters in the approved verification status. Which is not normal

- Reduce approving loans to small business purposes as defaulters are more

- Increase the interest rate where DTI is more