Lending Club Case Study

Problem Statement

You work for a consumer finance company which specializes in lending various types of loans to urban customers. When the company receives a loan application, the company must decide for loan approval based on the applicant's profile.

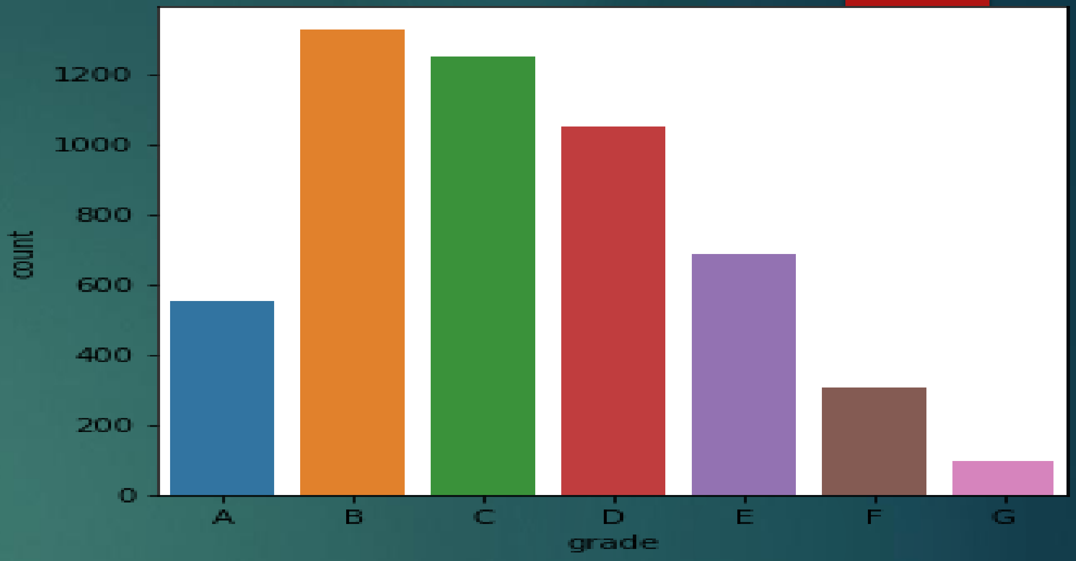
The aim is to identify patterns which indicate if a person is not likely to repay the loan i.e. be a defaulter which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

Analysis Approach

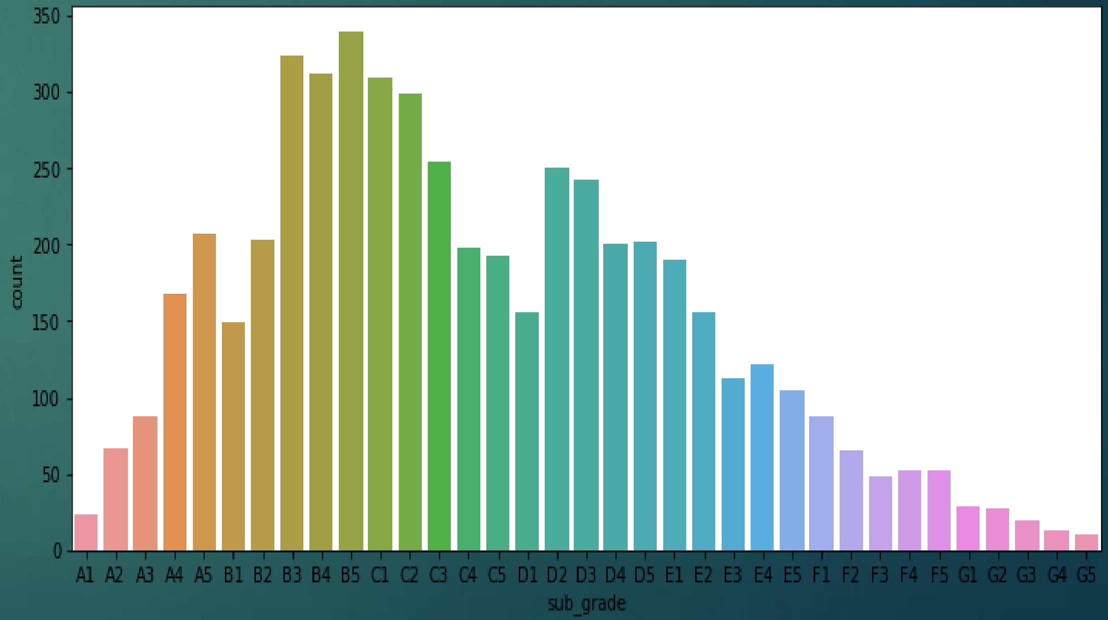
- Import the data
- Cleaning the data
- Univariate and Segmented Univariate Analysis
- Bivariate Analysis
- Conclusion

Exploratory Data Analysis

- Loans are paid off in order of the Grades with A being the best.
- Grade B and C have the probability of having more defaulters.

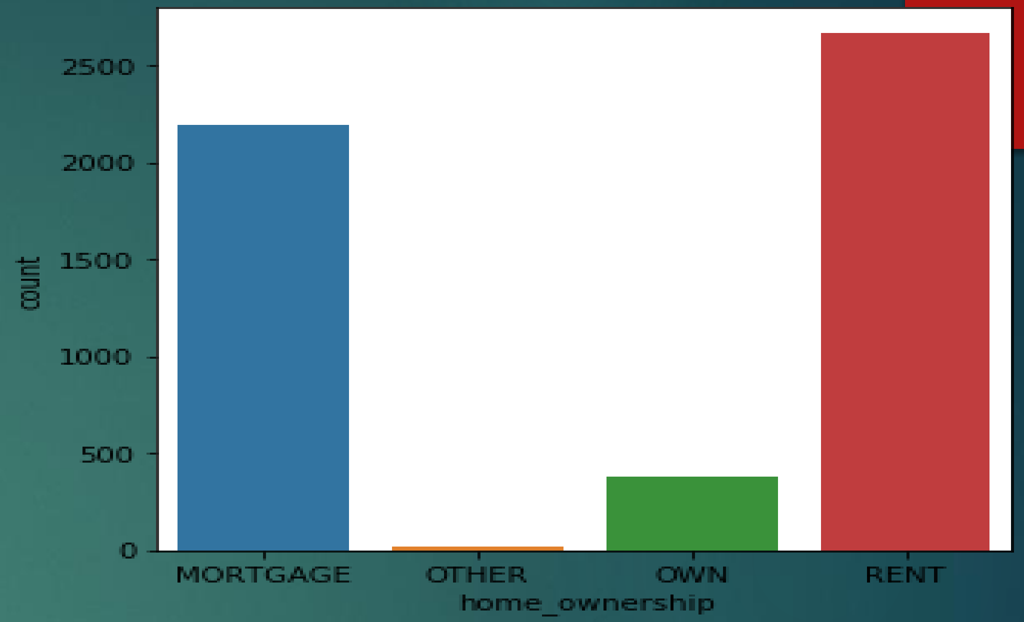


- Sub Grade shows exact same trend as Grades moves from B to G.
- Defaulters are more under Subgroup B onwards and gradually reduces towards F1.

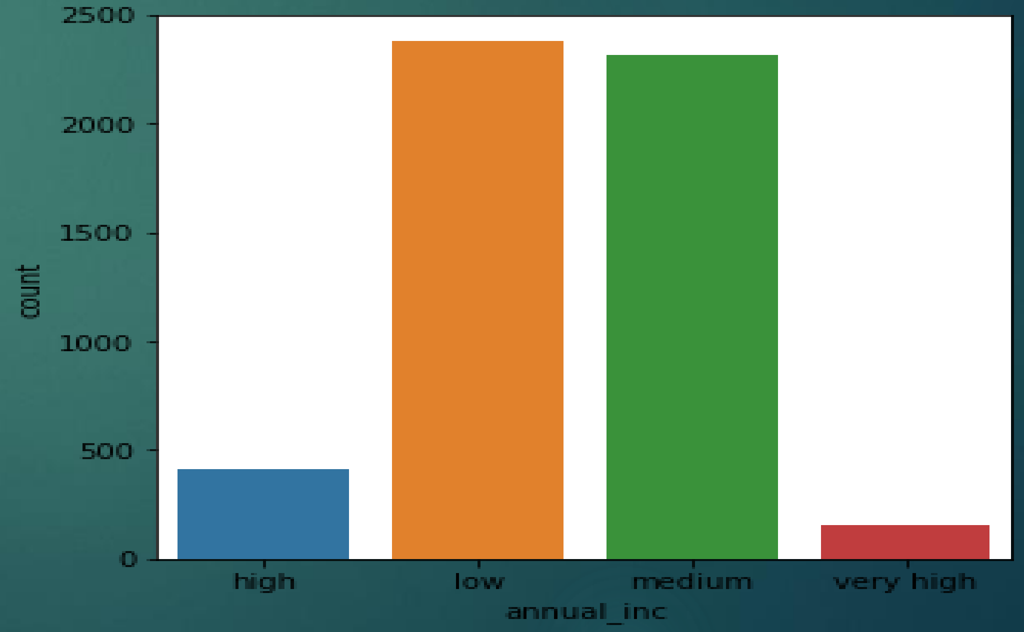


Exploratory Data Analysis

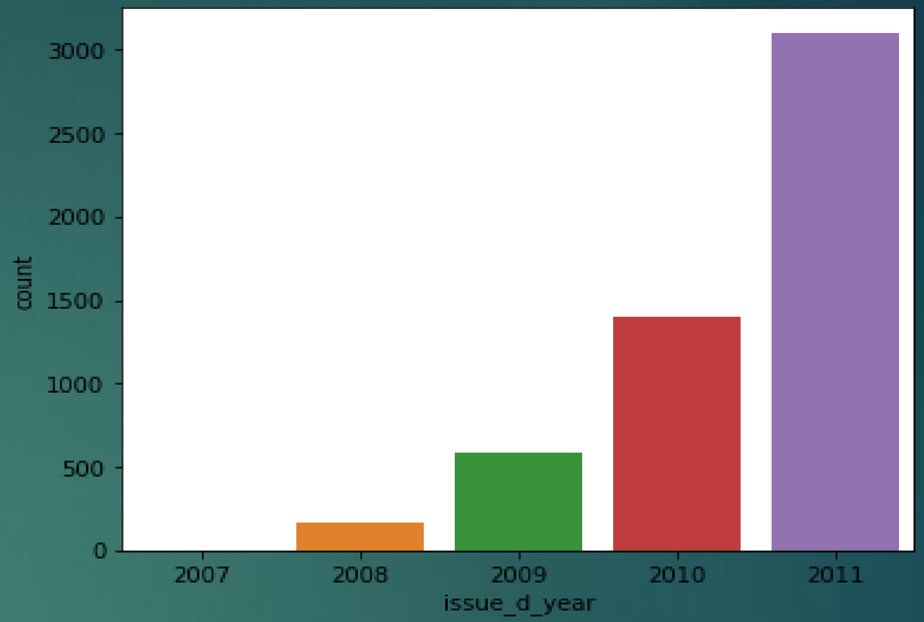
- home_ownership – is the status provided by the borrower during registration
- Renters and Mortgage have a higher probability of charge-off.



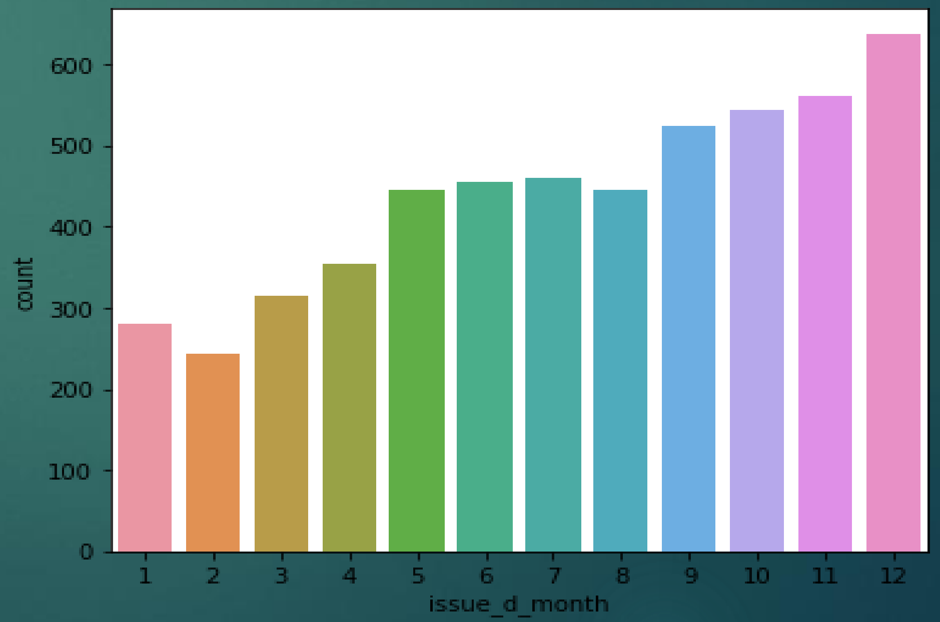
- Annual_inc is the annual income of the borrower.
- Borrowers with annual income less than 50000 default on their loans at much higher rate. Loan default decreases with higher income.



Exploratory Data Analysis



- 2011 is the year with most loans issued for both Fully Paid and Charged Off.

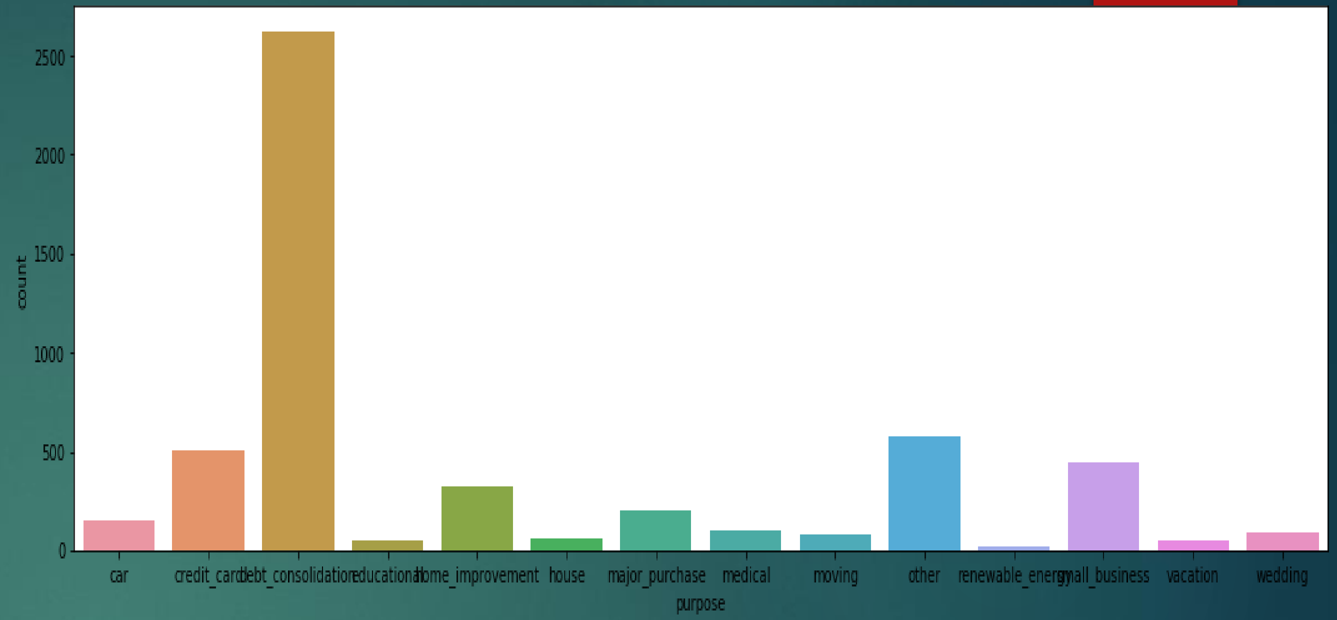


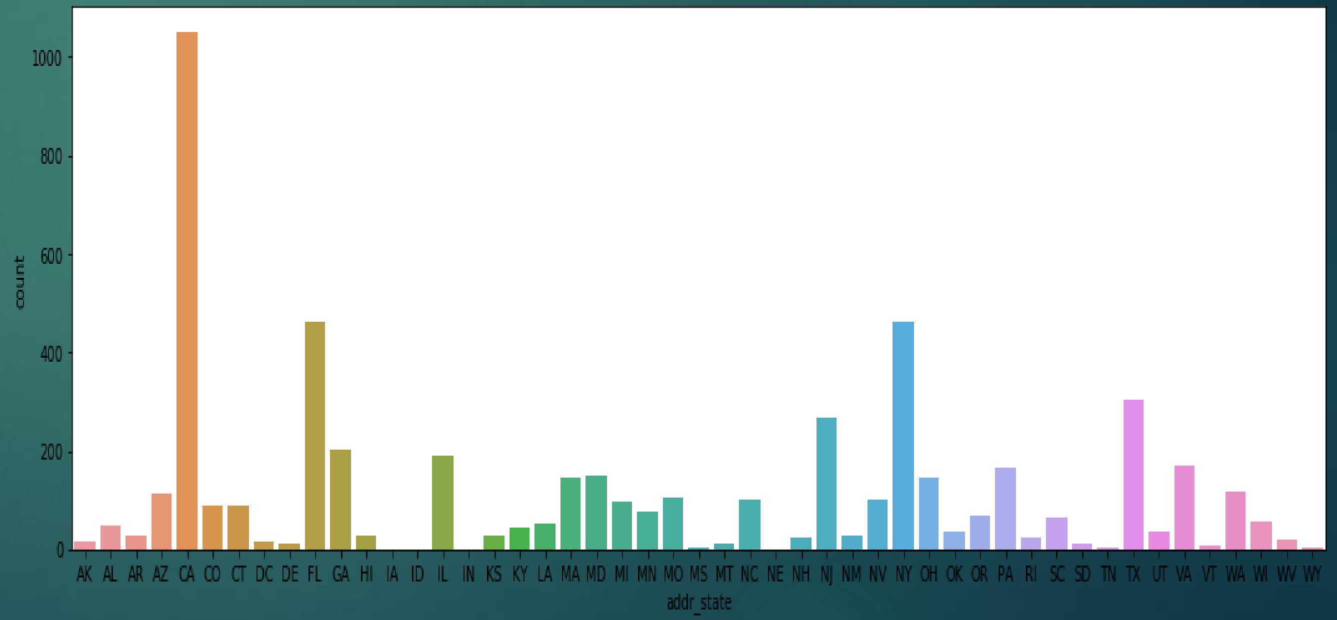
- Even the trend for the month in which the loan is issued is same for Fully Paid and Charged Off.
- Towards the end of year – most of the loans get issued.
- Not a significant factor to determine for Charged Off.

Exploratory Data Analysis

- Purpose is the borrowers stated reason for taking the loan.
- Reason debt consolidation and other have more defaulters.
- Lending company can look into purpose of taking loan and can further evaluate if the reason is other.

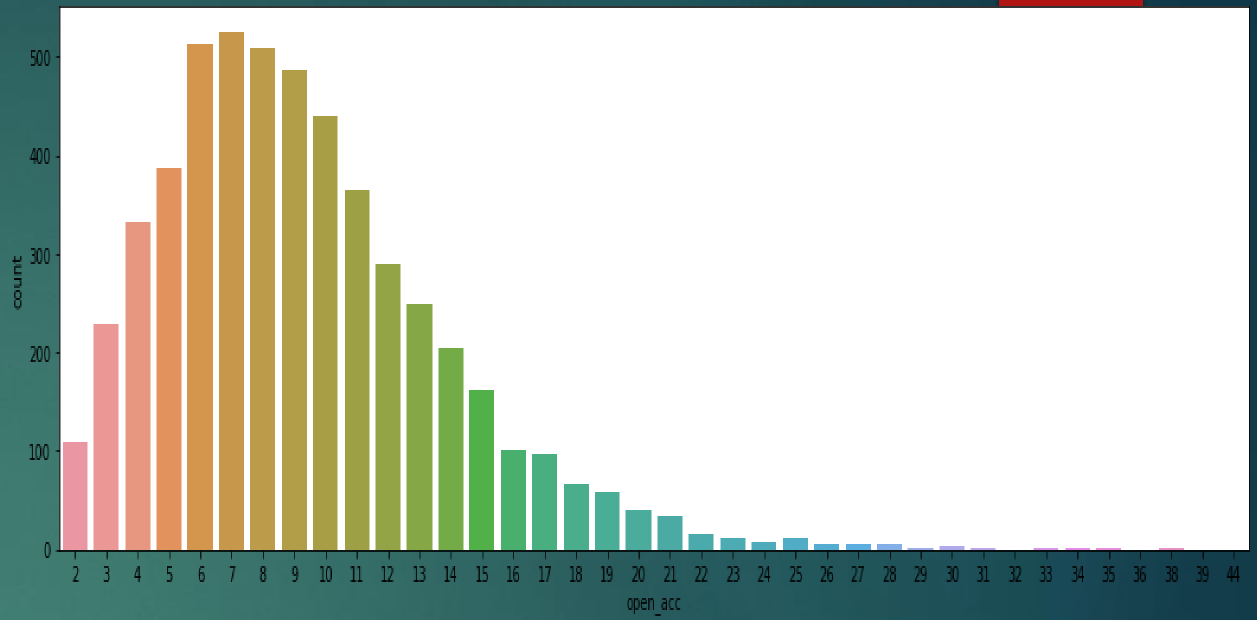
- Top five stated with most of the defaulters are California , Florida , New York, Texas and New Jersey
- These are the most populous states in USA as well.



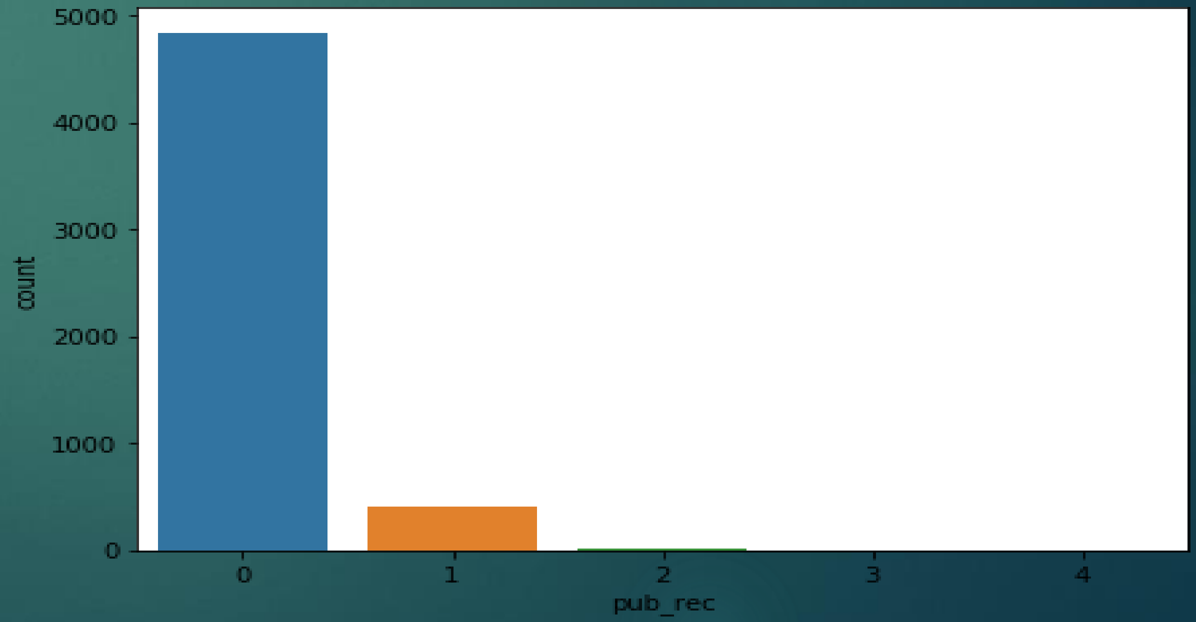


Exploratory Data Analysis

- The number of open credit lines in the borrower's credit file.
- Borrowers who have 3 or more credit lines open tends to be under defaulters .

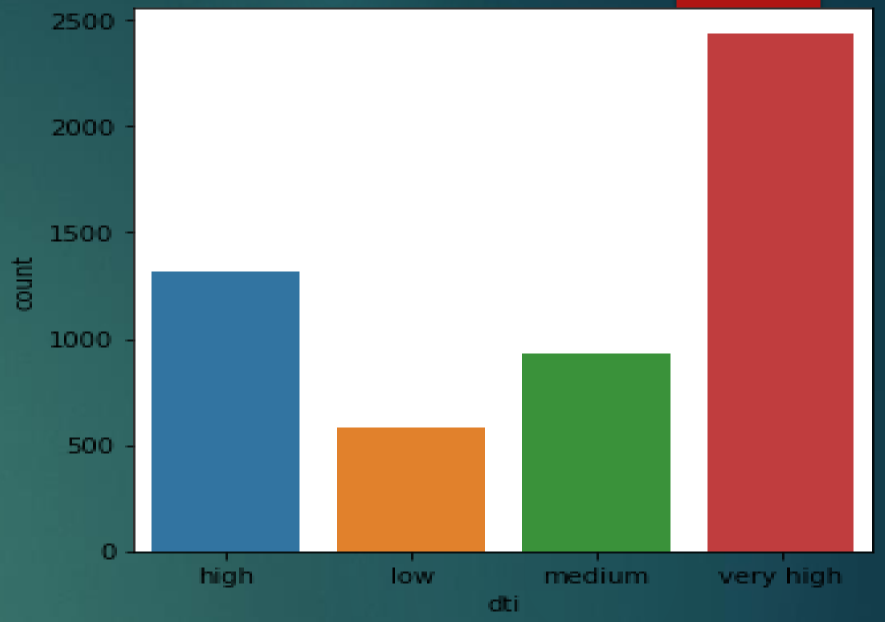


- Pub_rec is the number of derogatory public records against the borrower.
- More derogatory charges decreases the chances of getting the defaulters.

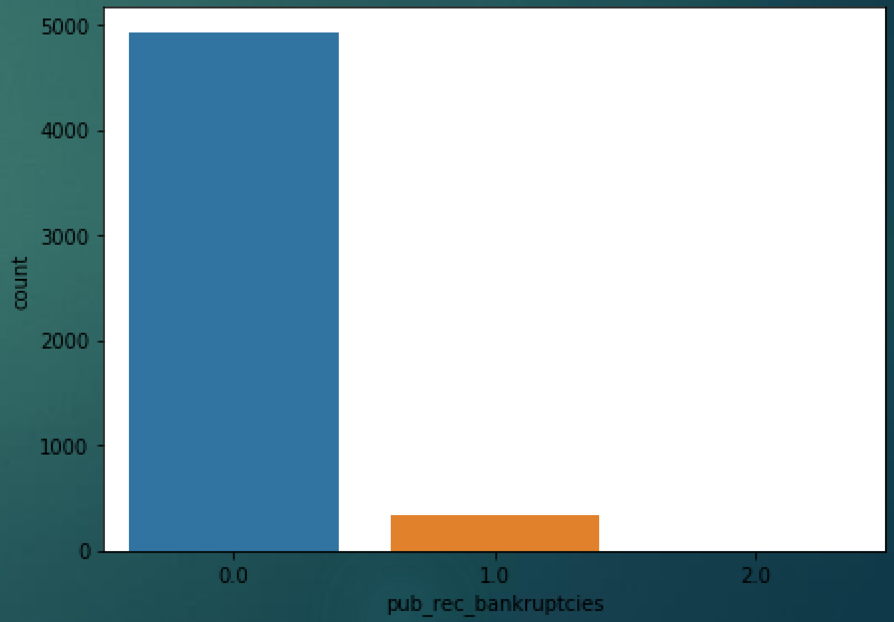


Exploratory Data Analysis

- dti is a ratio calculated using the borrower's total monthly debt payments on the total debt obligations, excluding mortgage and the requested LC loan, divided by the borrower's self-reported monthly income.
- Probability of borrowers which high and very high (more than 10) dti of becoming defaulters is high.

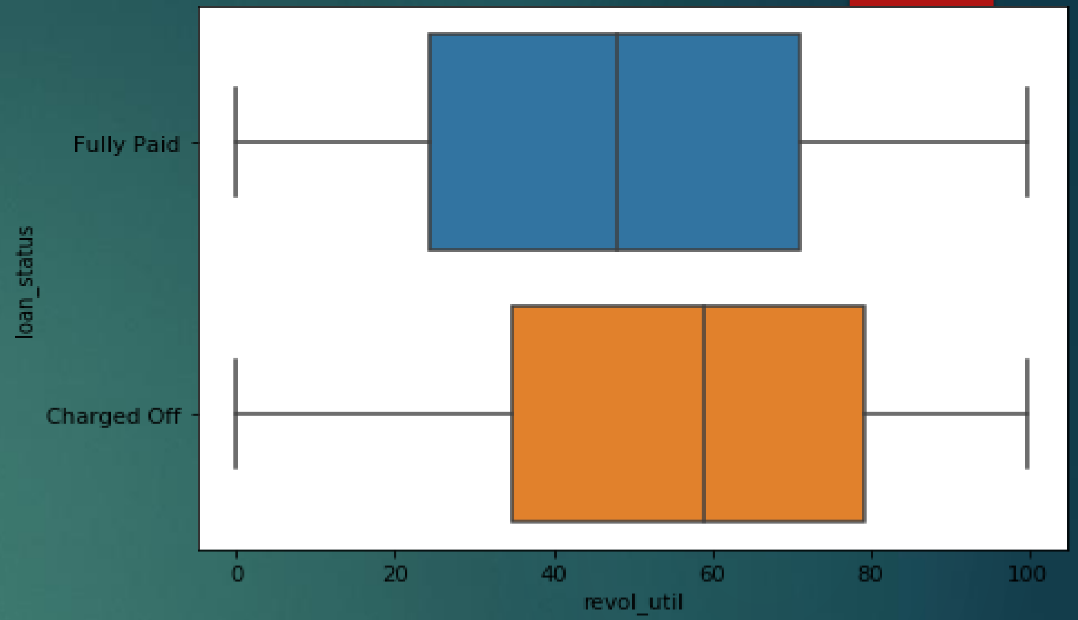


- Pub_rec_bankruptcies is number of public record bankruptcies
- Borrowers with 0 or 1 have the tendency of defaulters.

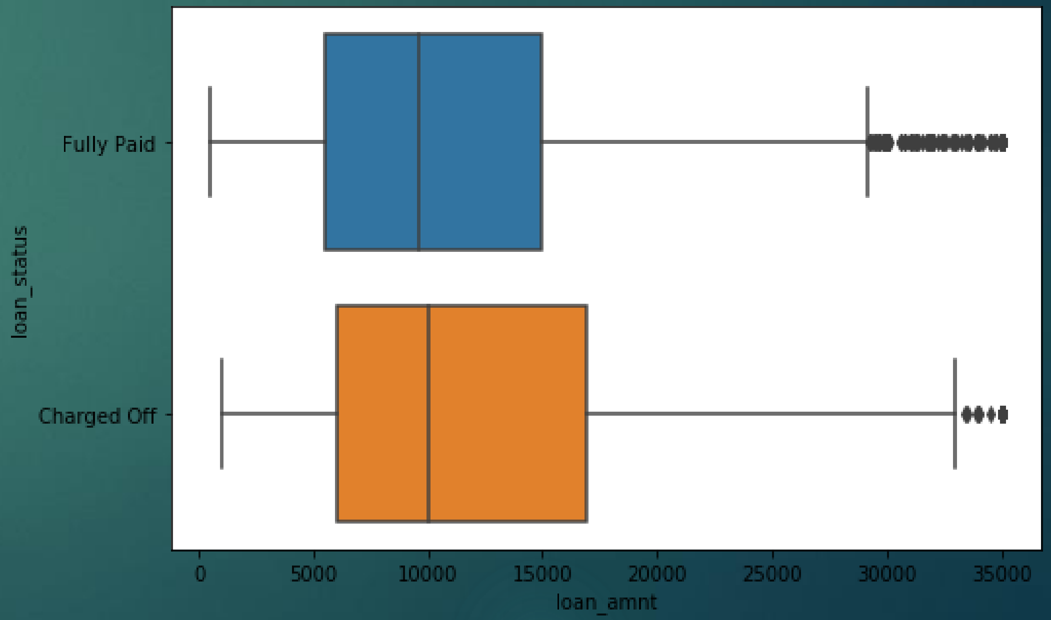


Exploratory Data Analysis

- ➢ Revolving line utilization rate, or the amount of credit the borrower is using relative to all available revolving credit.
- ➢ Charged Off seems to have probability of having more revol_util.

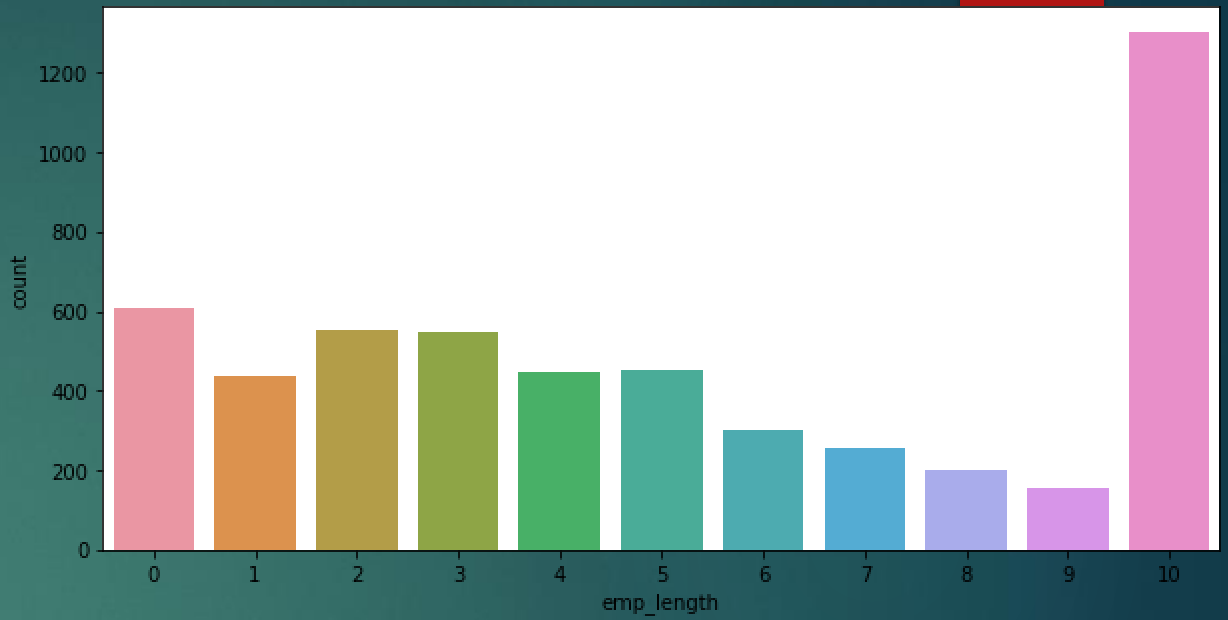


- ➢ The listed amount of the loan applied for by the borrower.
- ➢ Charged off are shifted towards higher average loan amount request.



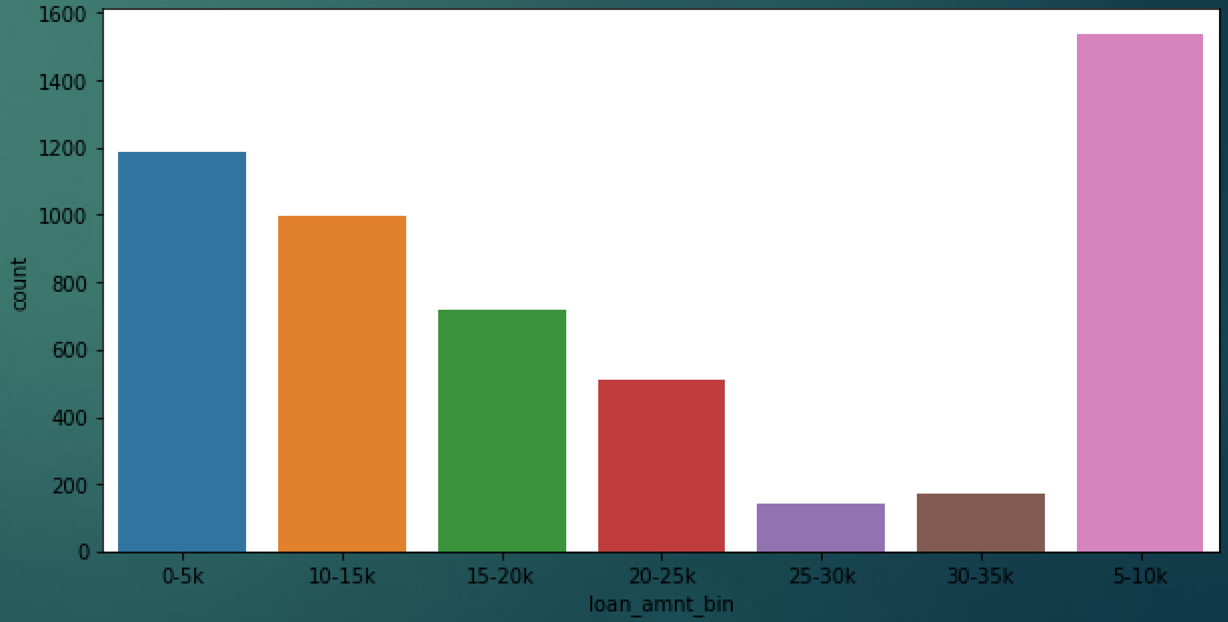
Exploratory Data Analysis

➢ Borrowers with 10 years of more employment length tends to default more than those with less years of experience.

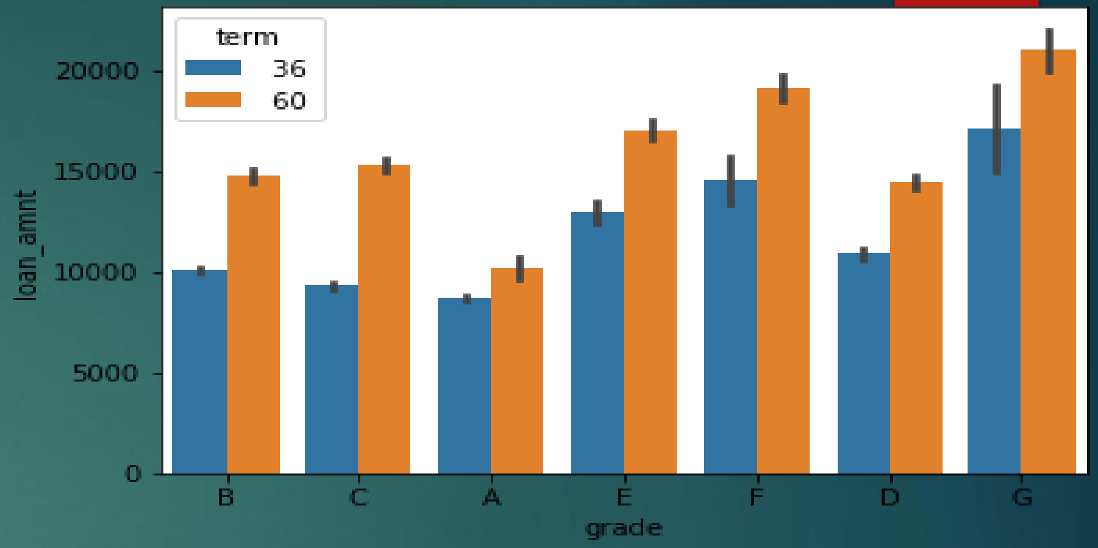


➢ Loan amount is the actual loan applied by the borrower.

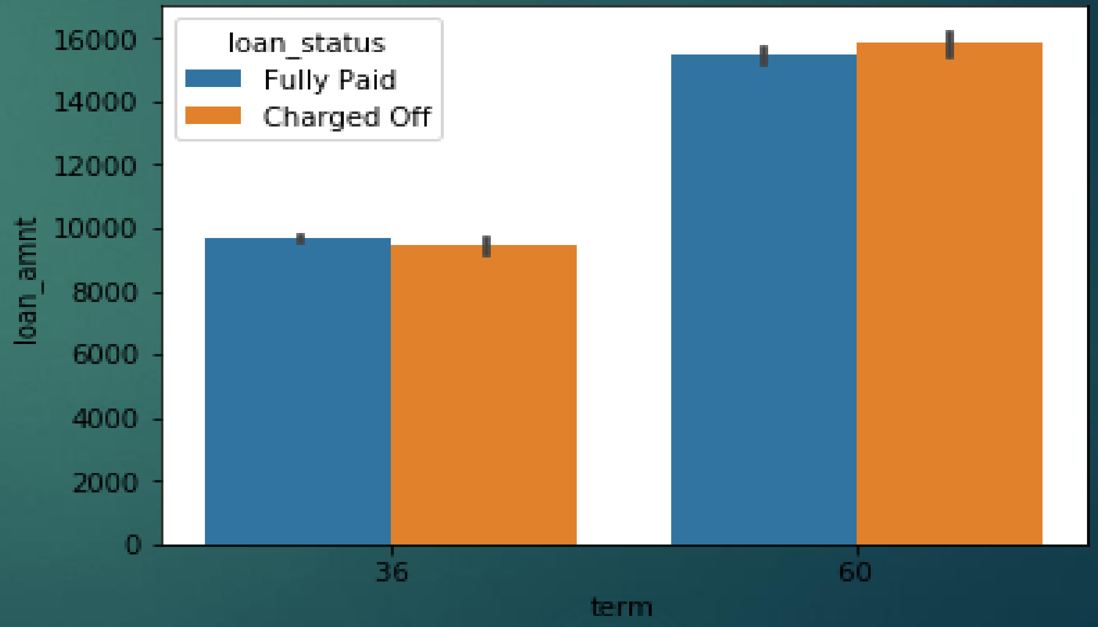
➢ Probability of borrowers defaulting have the loan amount of 5-10K .



Exploratory Data Analysis



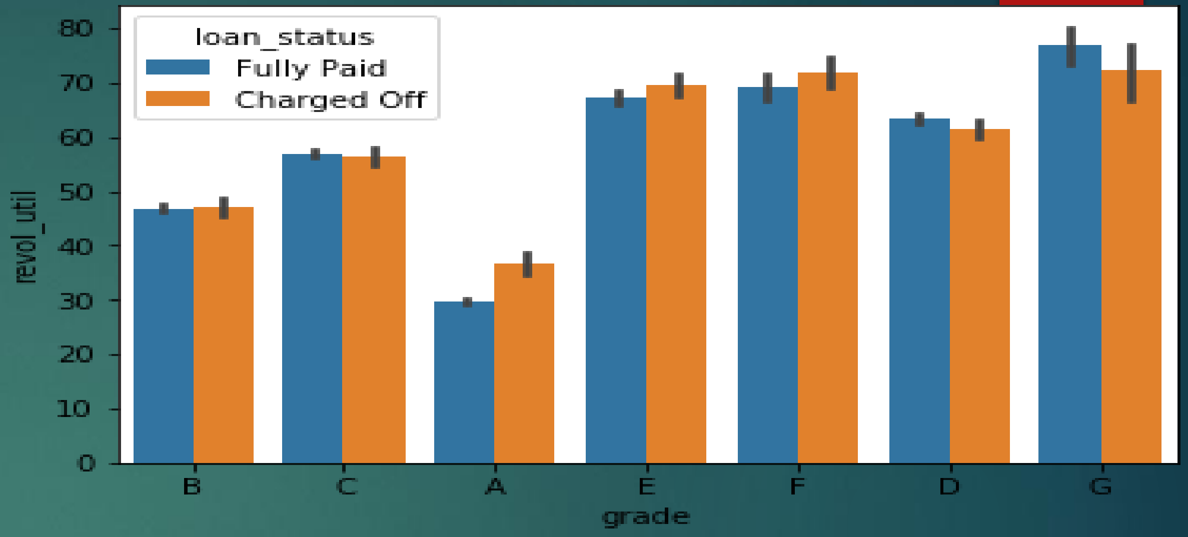
- ➢ Higher loan amount are associated with lower grade for longer terms



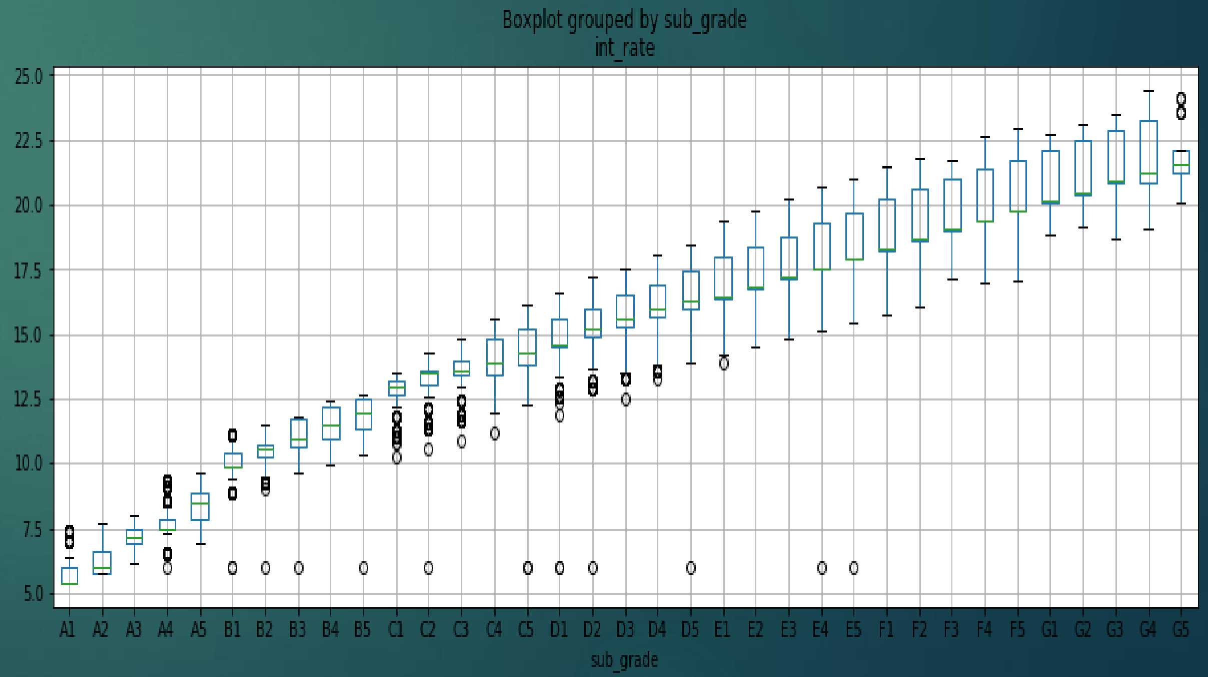
- ➢ Higher loan amount are associated with longer terms and see higher Charge Offs.

Exploratory Data Analysis

- ➢ revol_util and grade(and therefore int_rate) are correlated in some way. The revol_util is positivly correlated to the grade. As the grade goes from A to E the revol_util also increases.
- ➢ This may be because higer loan amounts are associated with higher grades.

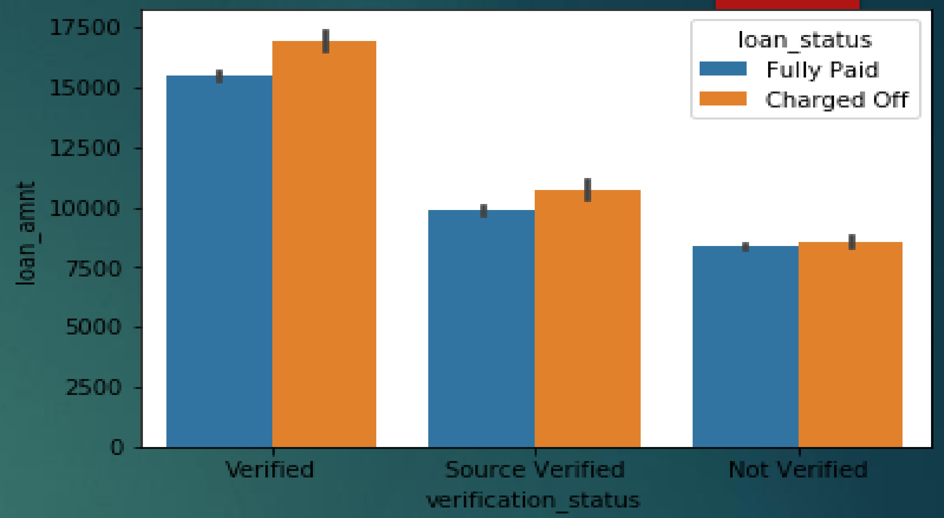


- ➢ Interest rates varies directly with the subgrade. Larger or worst the sub grade, higher are the rate of interest for the loan.

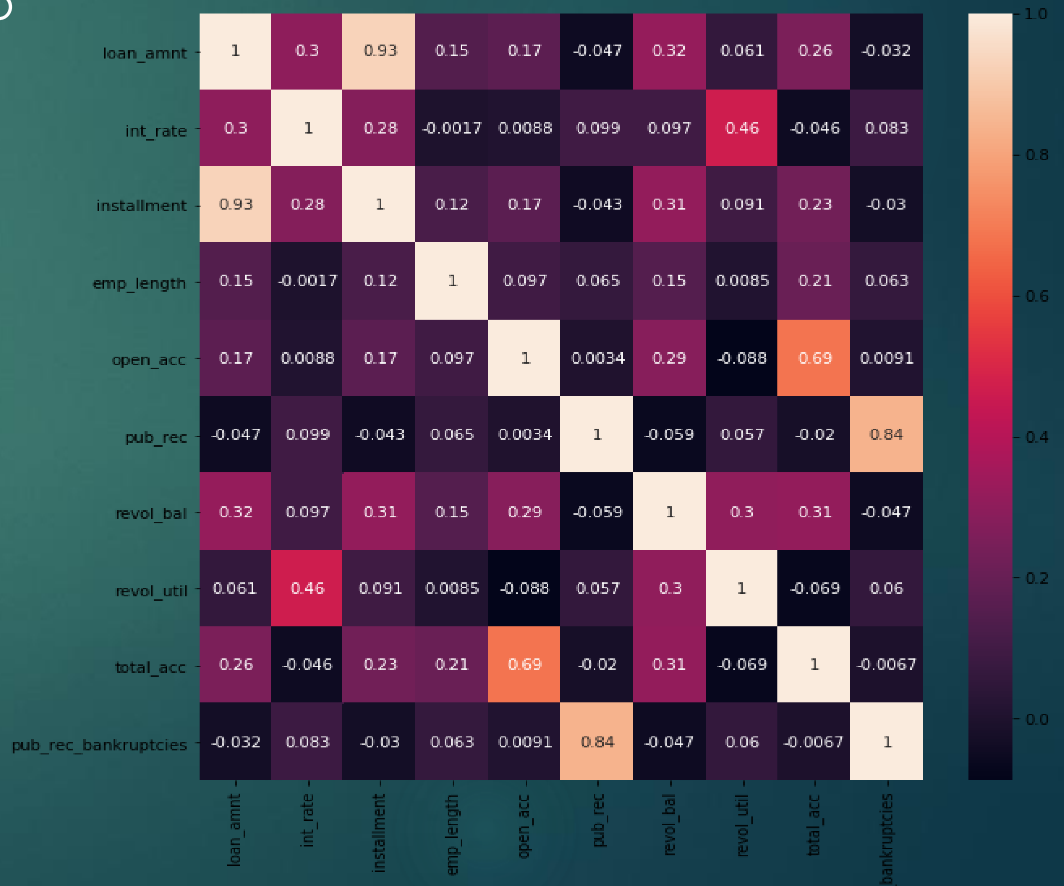


Exploratory Data Analysis

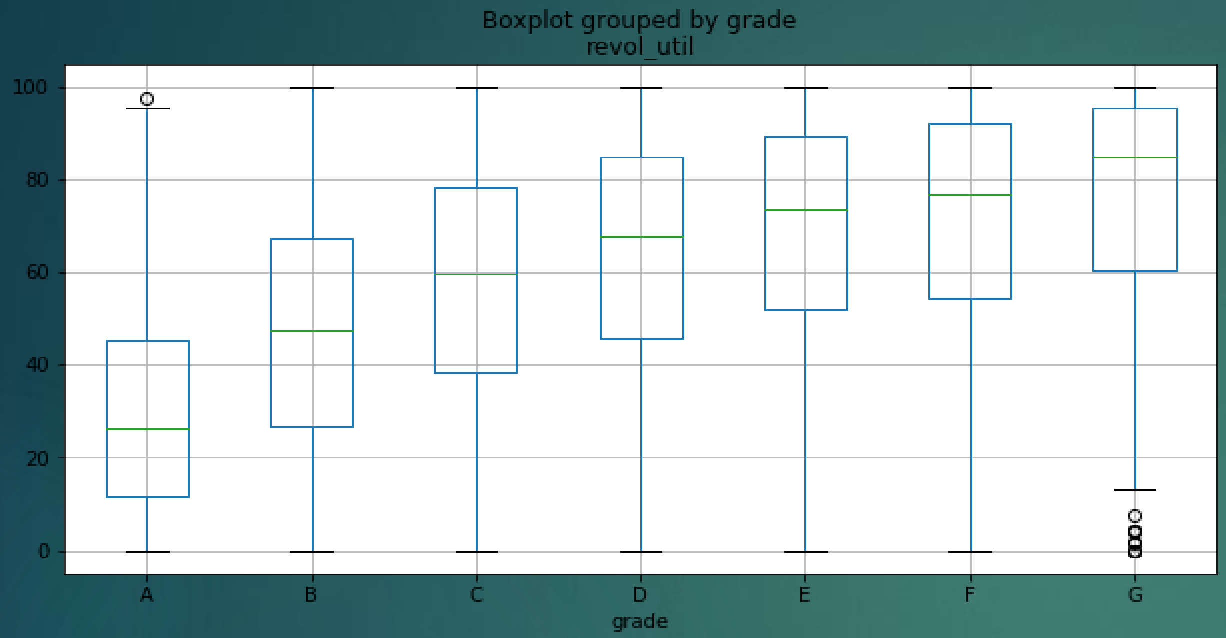
- verification status indicates if income was verified by LC, not verified, or if the income source was verified
- verification_status with value as Verified seems to have more probability of getting Charged_off

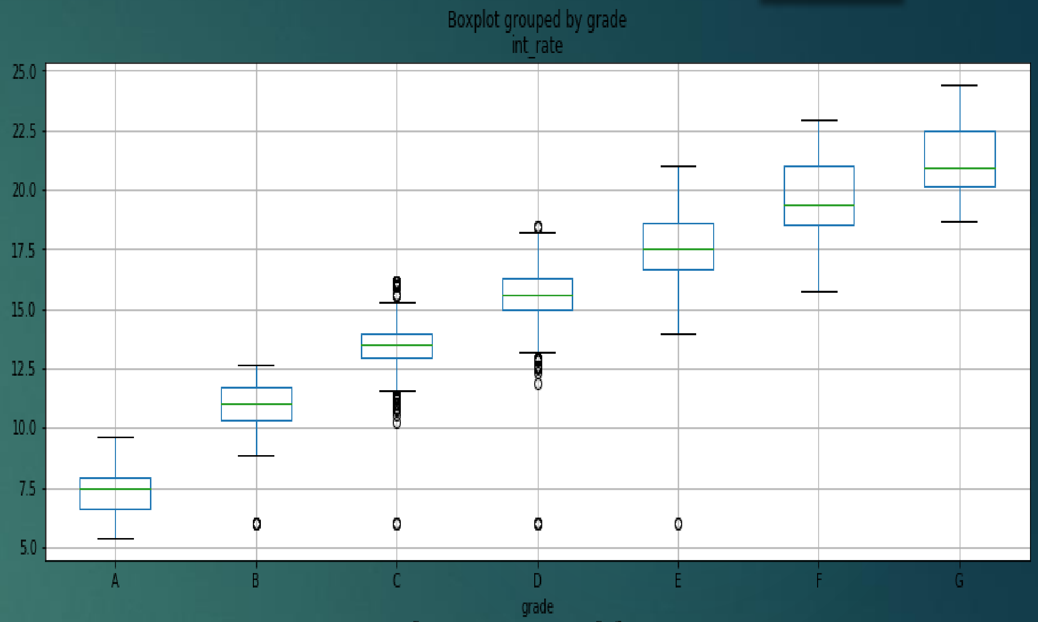


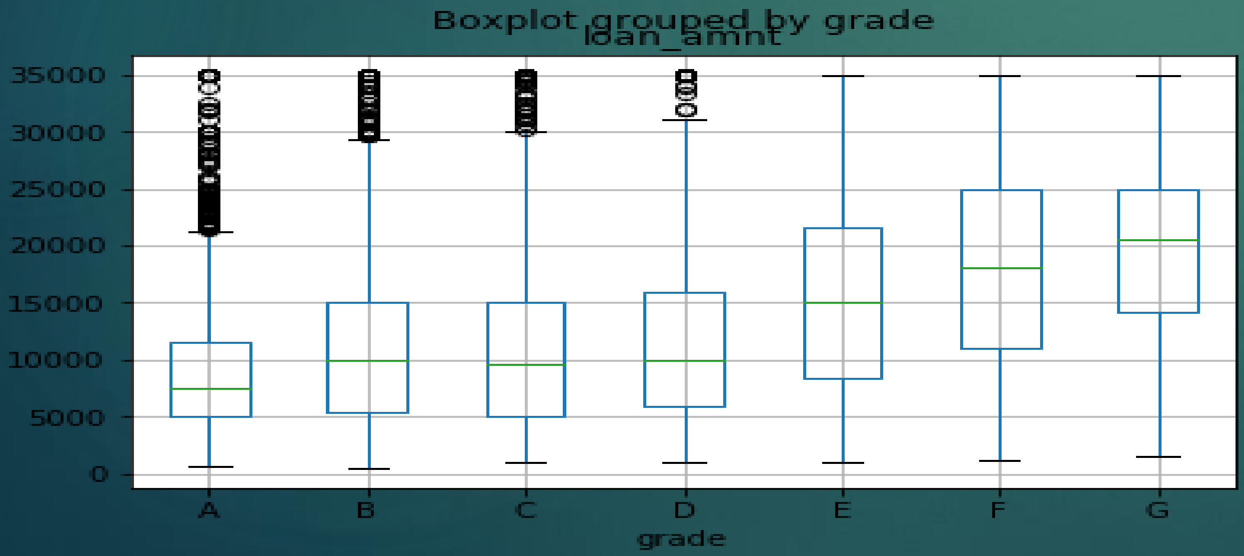
- Heat map showing few variables with the correlation matrix.

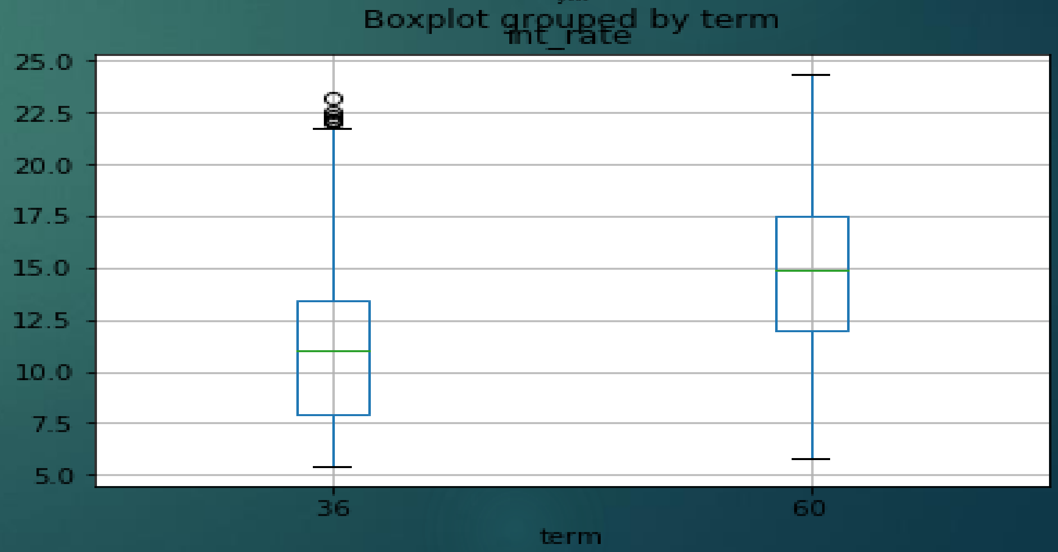


Few more …


Boxplot grouped by grade
revol_util
100
80
60
40
20
0
A
B
C
D
E
F
G
grade


Boxplot grouped by grade
int_rate
25.0
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0
A
B
C
D
E
F
G
grade


Boxplot grouped by grade
loan_amnt
35000
30000
25000
20000
15000
10000
5000
0
A
B
C
D
E
F
G
grade


Boxplot grouped by term
int_rate
25.0
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0
36
60
term

Conclusion

Below are the variables which are strong indicators of looking into while approving the loan:

Grade/Sub Grade/Interest Rate
Home ownership status
Loan Purpose
Loan Amount
Annual Income
Verified Income Status
Revolving Utilization
Term of the loan
Employment length
Installment Amount
Verified Income Status conflicts with its purpose i.e. those who have the income status verified shows the result opposite of what we would expect.