

Lending Club Case Study

By:
Santhosh Kumari Kage
Deepesh Rawat

Agenda

- Problem statement
- Analysis Summary
- Univariate Analysis & Conclusions
- Bivariate Analysis & Conclusions
- Multivariate Analysis & Conclusion

Problem Statement

Identify the features or variables that will help process the loan applications that will reduce the risk of loans not defaulting.

Analysis Summary

Provides us with insight into what has been done and the considerations to derive certain conclusions.

- Post understanding the data and performing the required clean-up operations as per EDA we were able to get the columns of interest for the target variable.
- Columns of interest will help in determining if loans will default or not. Below are the details for same:

Categorical columns

"term","grade","sub_grade","emp_length","home_ownership","verification_status","purpose","addr_state","pub_rec_bankruptcies","loan_range","rate_interest","annual_inc_range","issue_d_month","issue_d_year","sub_grade_rank"

Continuous Columns

"installment","mths_since_last_delinq","dti"

Target Columns : which will determine where a person will pay the loan or will default on it.

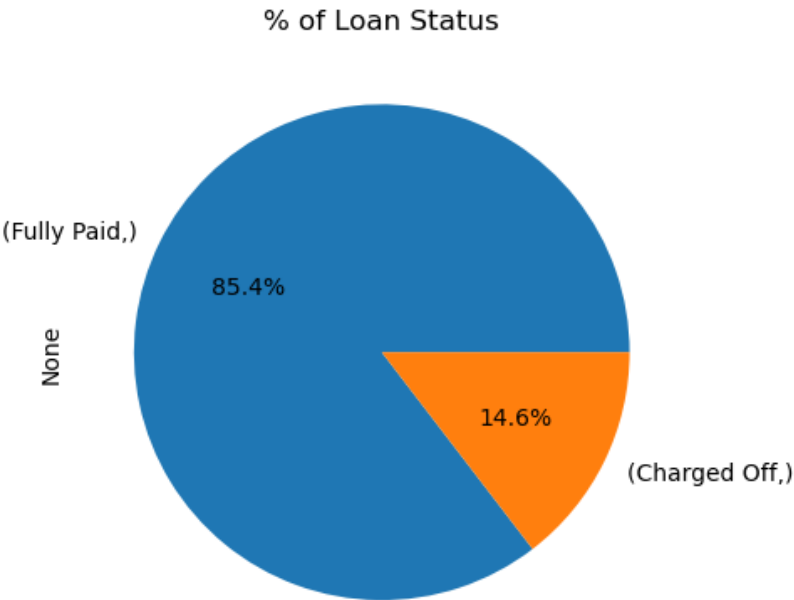
"loan_status"

- Performed various analysis (univariate, Bivariate, Multivariate) against these columns of interest and derived conclusions

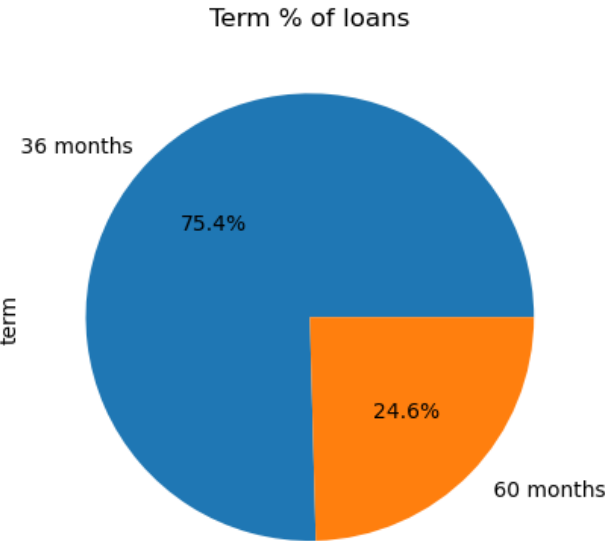
Univariate Analysis & Conclusion

Univariate Analysis & Conclusions

85% of loans are paid in full vs 25% who have defaulted

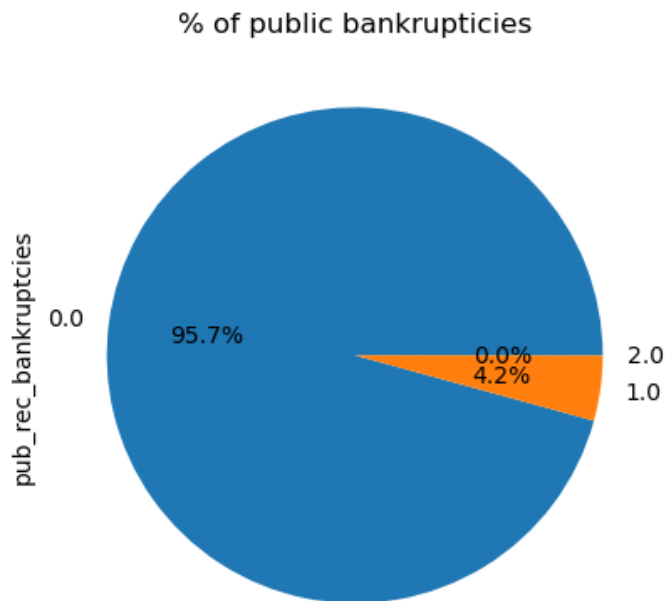


75% of applicants prefer 36 months of tenure for their loan

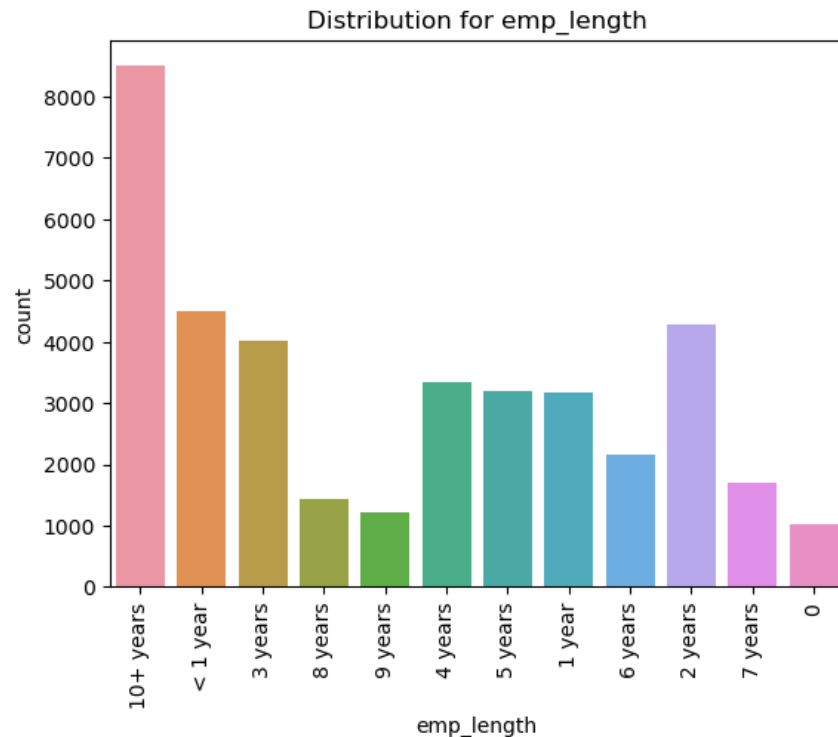


Univariate Analysis & Conclusions

Around 4.2 percent of users have at least 1 bank bankruptcies cases registered against them

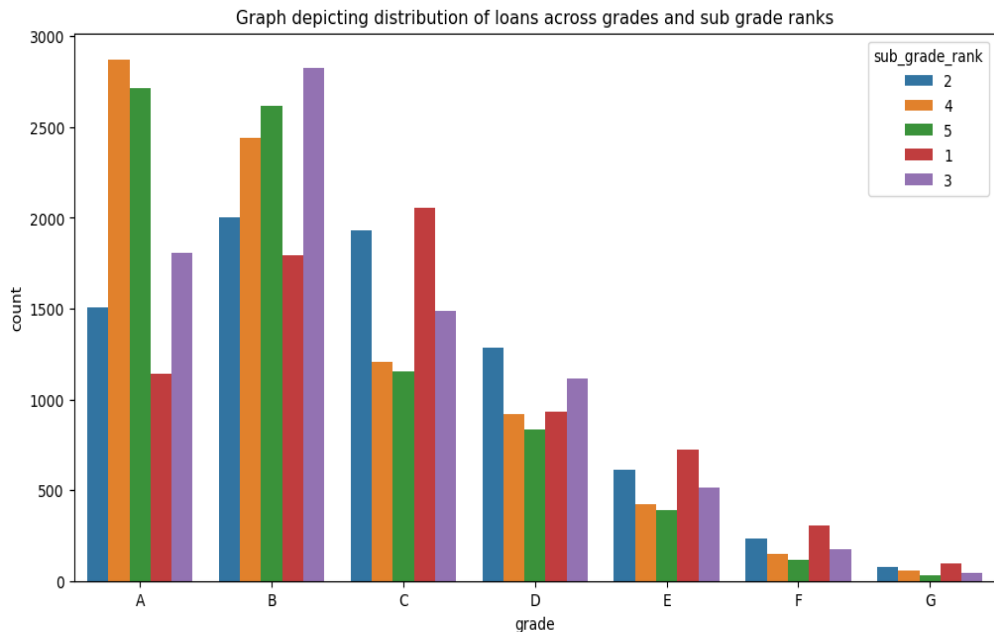
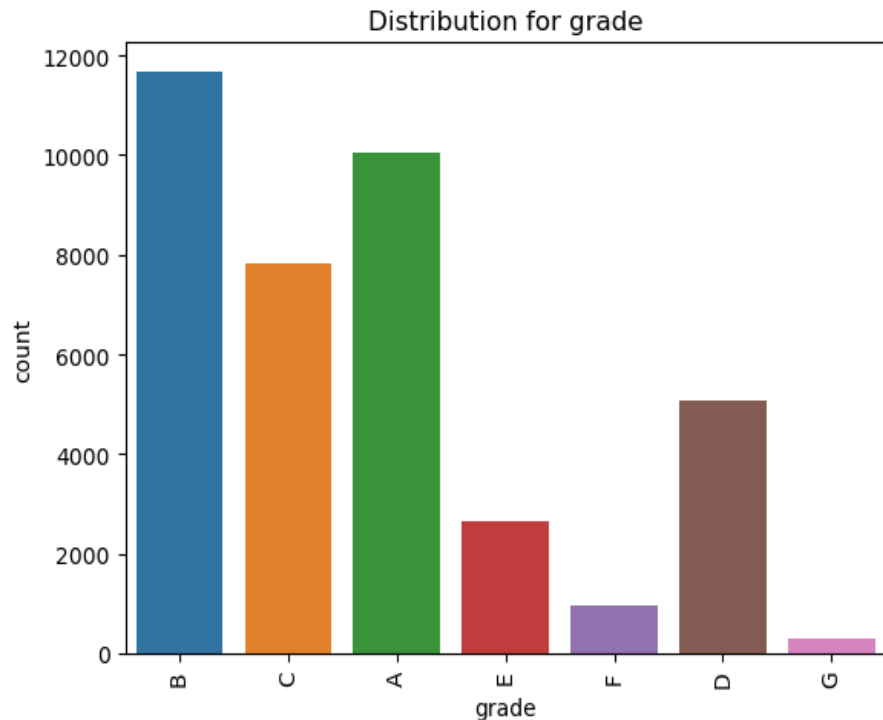


The major loan applicants are from employees who have more than 10 years of exp.



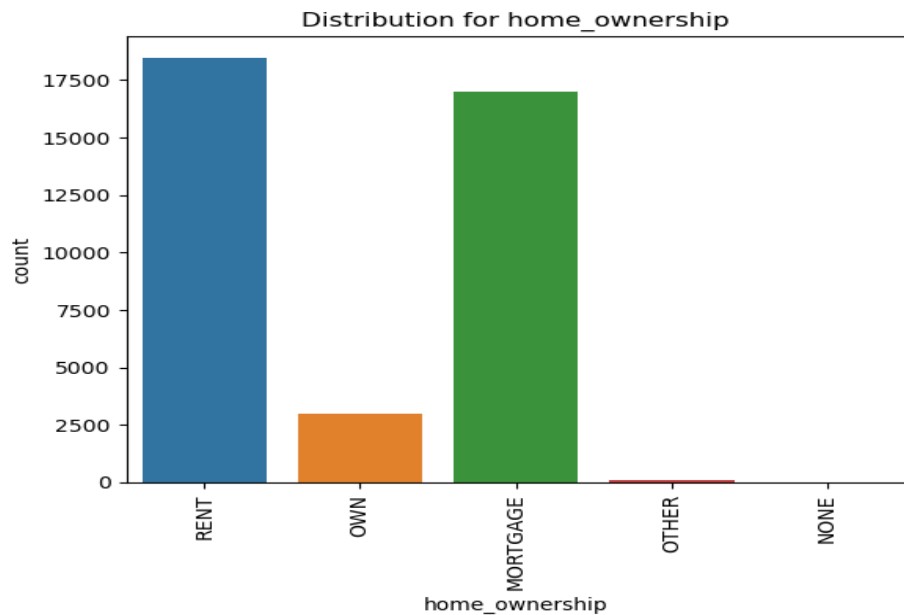
Univariate Analysis & Conclusions

Applicants are mostly categorized on grade B (B3 major subgrade) followed by grade A (A4 major subgrade) based on the base rate and risk.

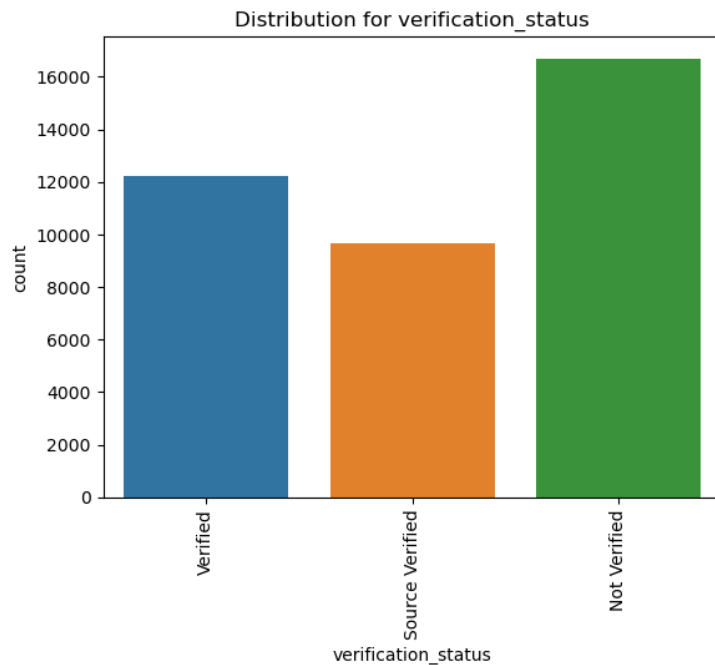


Univariate Analysis & Conclusions

People who rent/ mortgage out the homes are the major applicants. The applicants who are homeowners are the least.

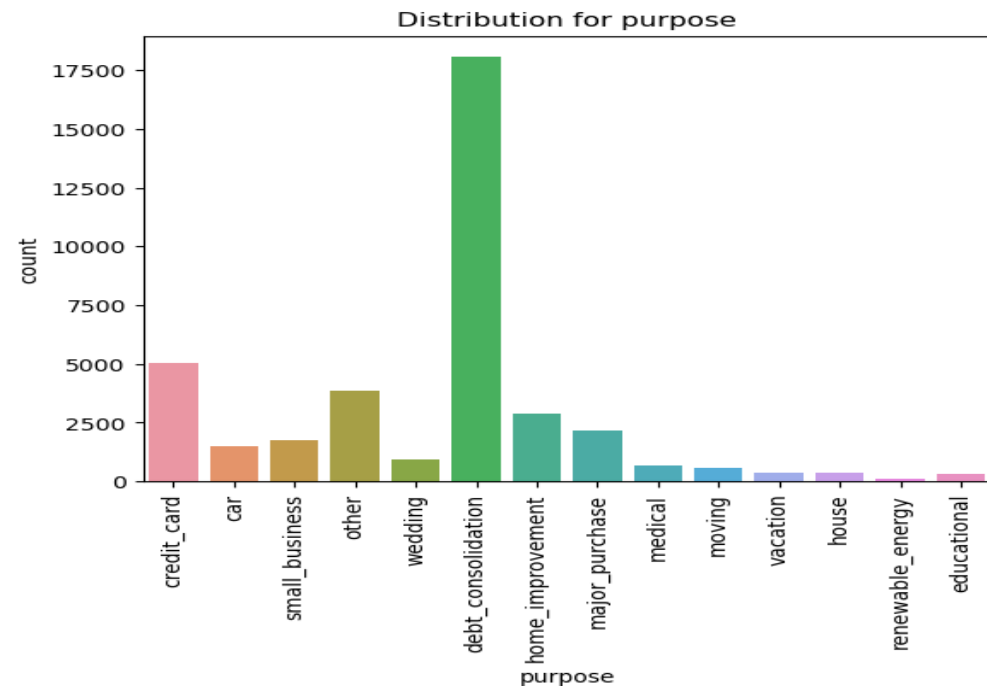


Most of the applicants are not verified.

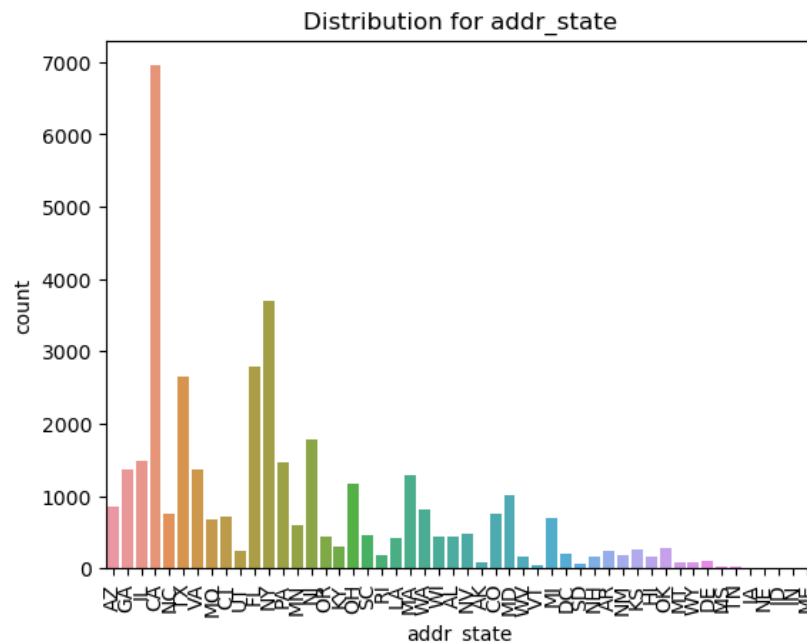


Univariate Analysis & Conclusions

Applicants usually take loans for debt consolidation.

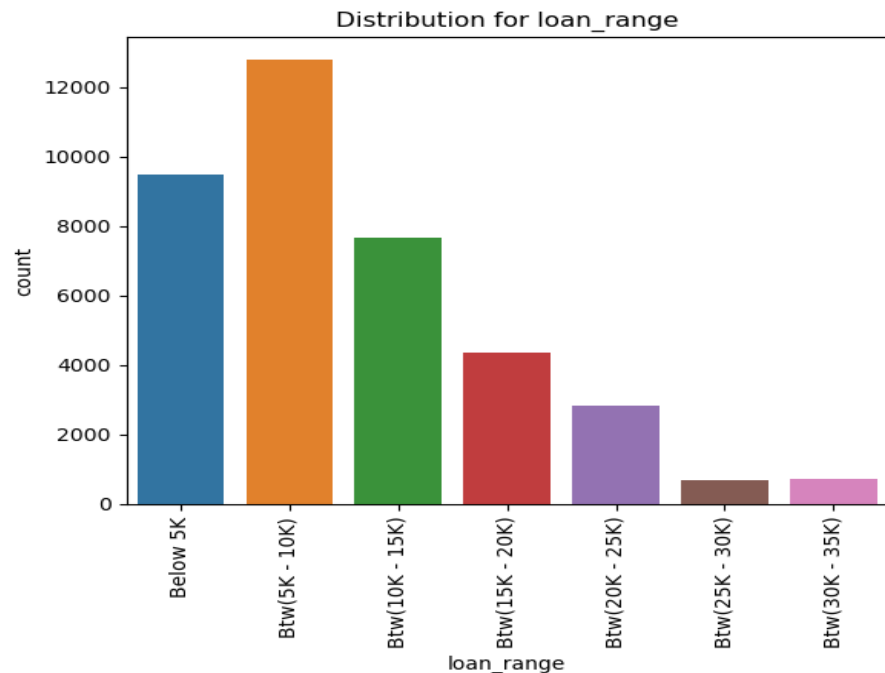


Most of the applications are from CA address.

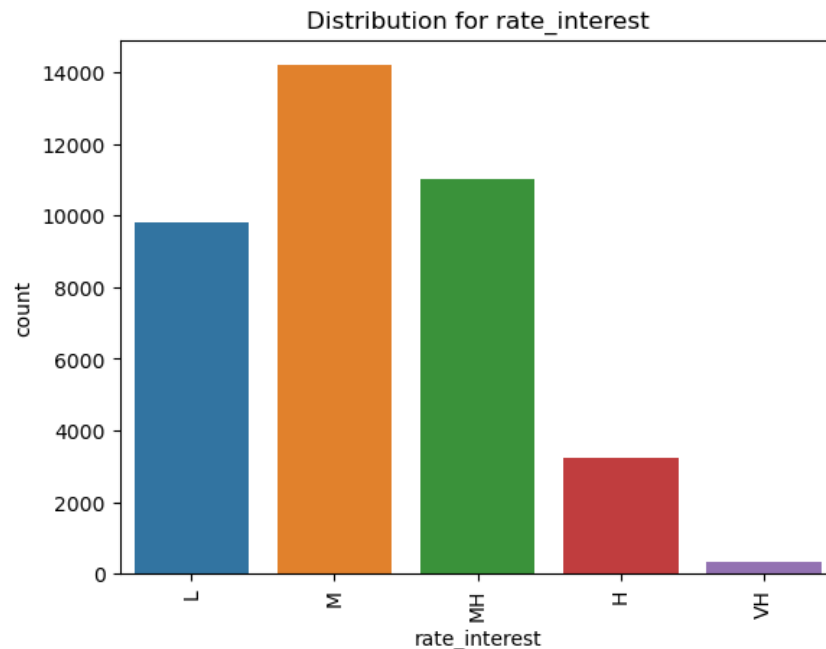


Univariate Analysis & Conclusions

Applicants are mostly looking for loans between 5K to 10K.



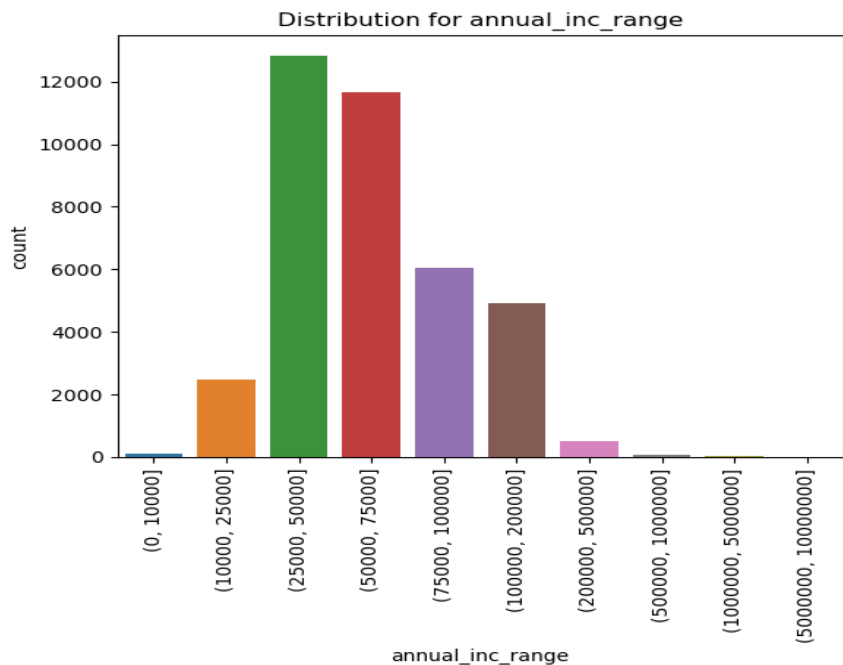
Most of the loans have an interest rate of Medium i.e., 9 – 13%



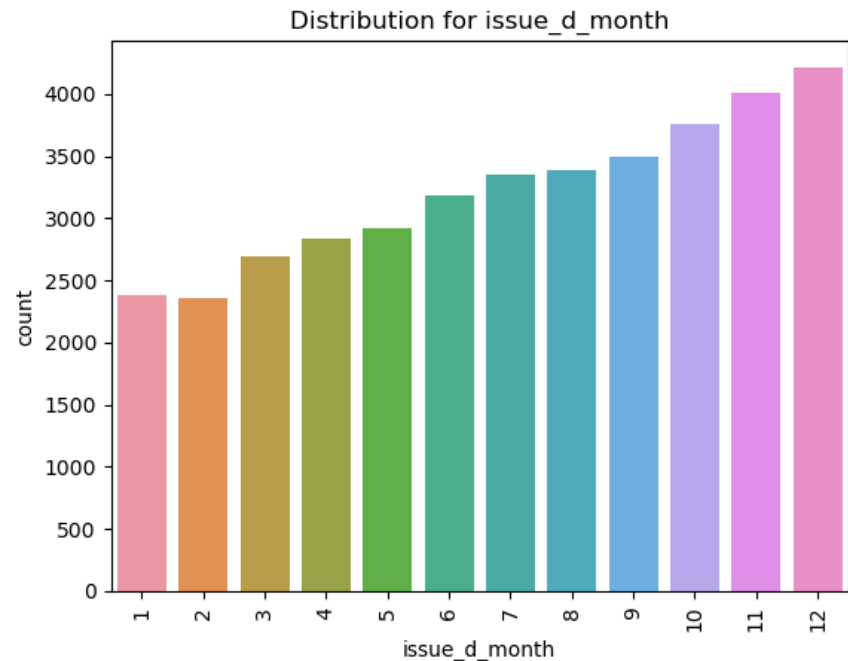
L - low - 5 to 9
M - Medium- 9 to 13
MH - medium high - 13 to 17
H - high - 17 to 21
VH - very high - 21 to 25

Univariate Analysis & Conclusions

Most of the loan applicants have an annual income of 25K to 50K



Most of the loan as issued in the month of Dec.
Probably the year end people want to pay/close their debts.

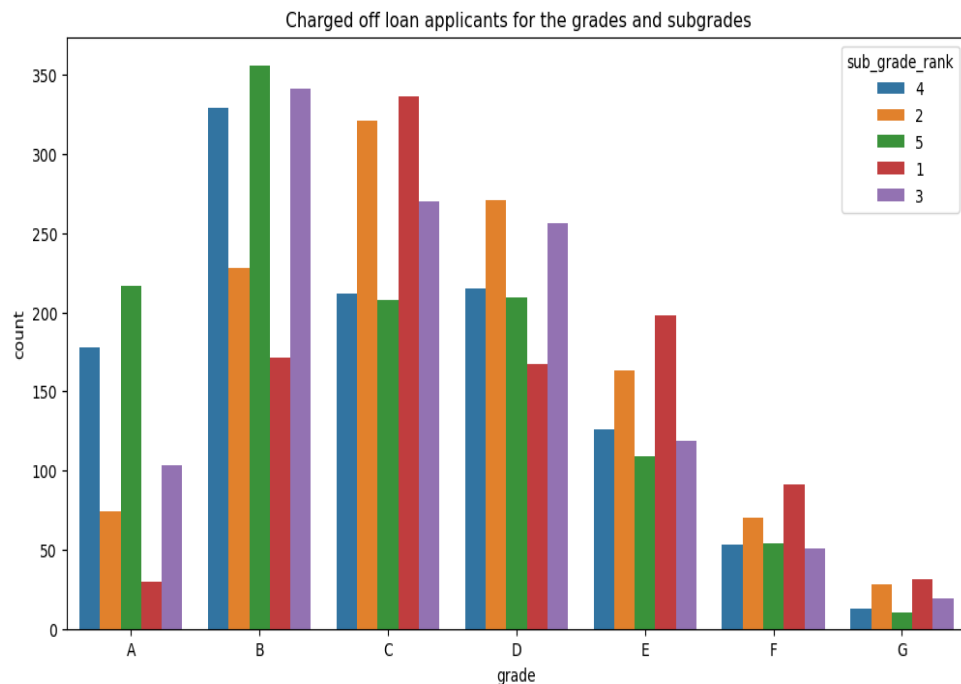
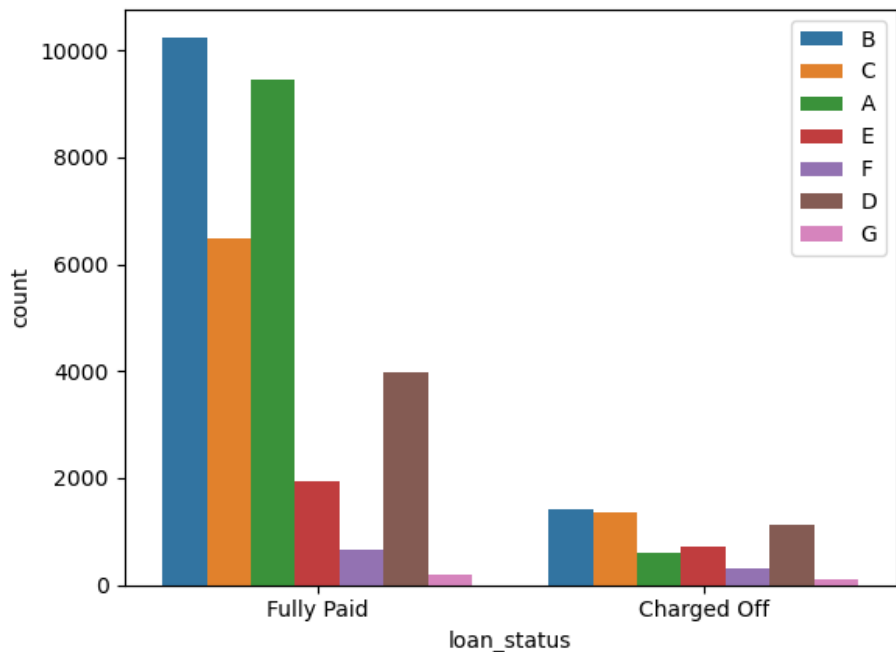


Bivariate Analysis & Conclusions

Bivariate Analysis & Conclusions

Loan applicants with a grade of B5 have the highest no of charged-off accounts followed by B3 and B4 in grade B. Followed by grades C - C1 and C2.

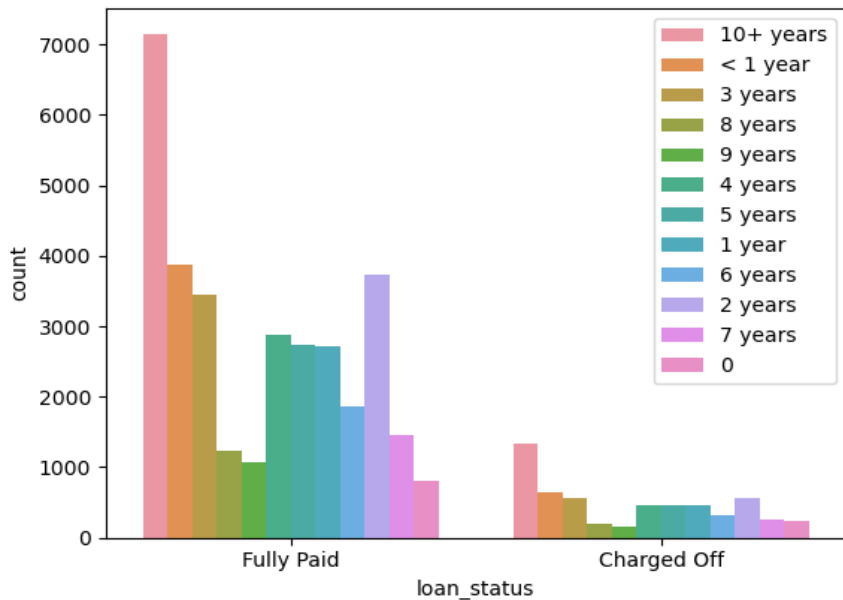
So loan applicants with grades B and C are more likely to default.



Bivariate Analysis & Conclusions

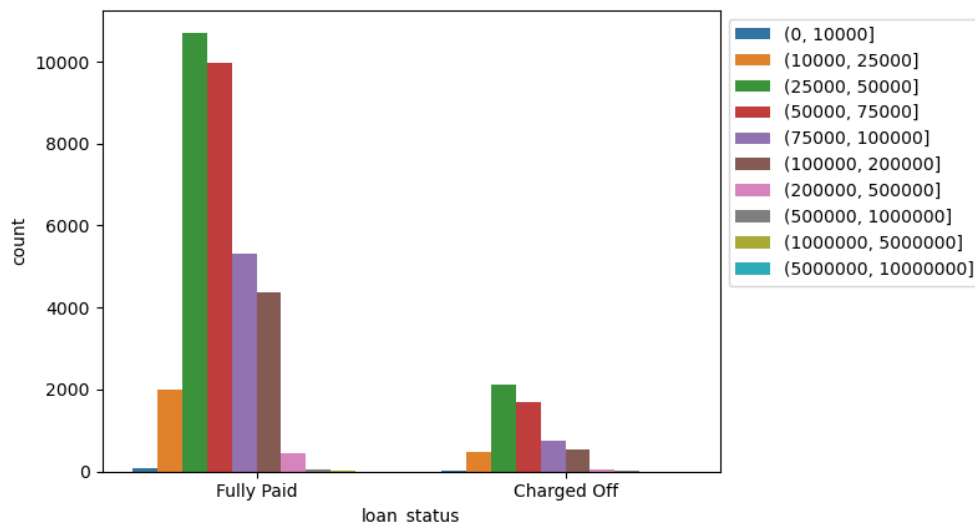
Employees with more than 10 years of experience are likely to have more loan defaults compared to others.

emp_length vs loan status



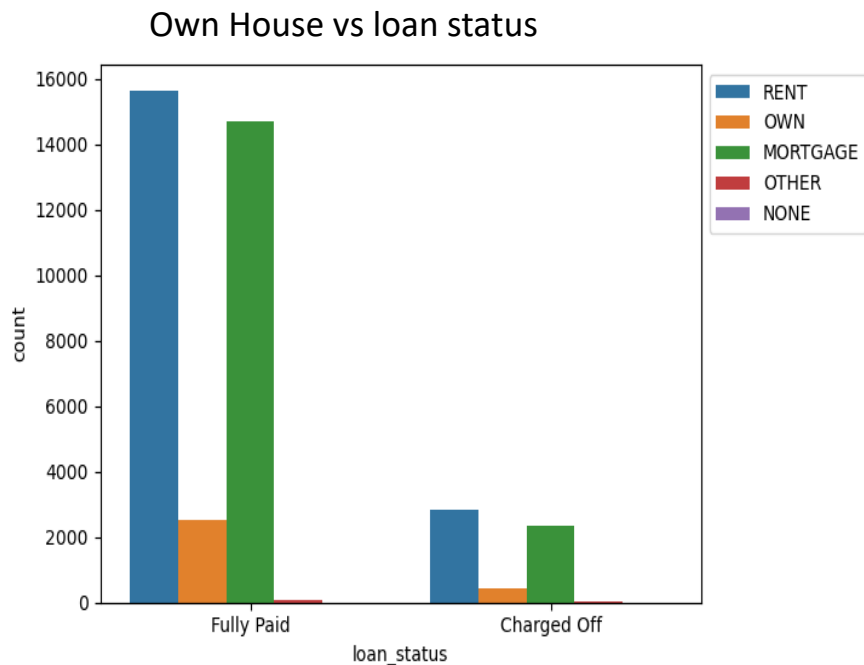
Loan applications having annual income between 25K to 50K are likely to have more loan defaults compared to others.

Annual income vs loan status

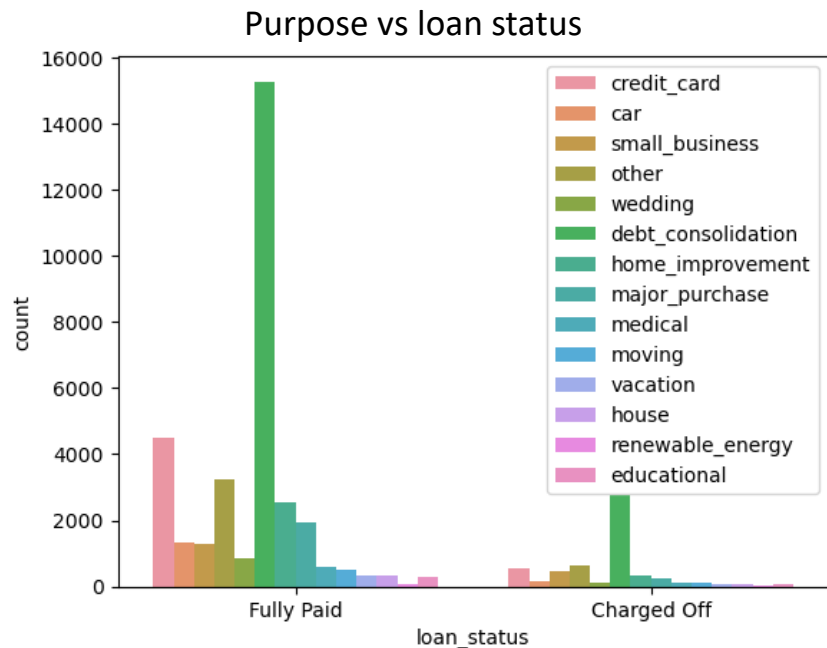


Bivariate Analysis & Conclusions

Loan applications who don't own a house are more likely to default on a loan.



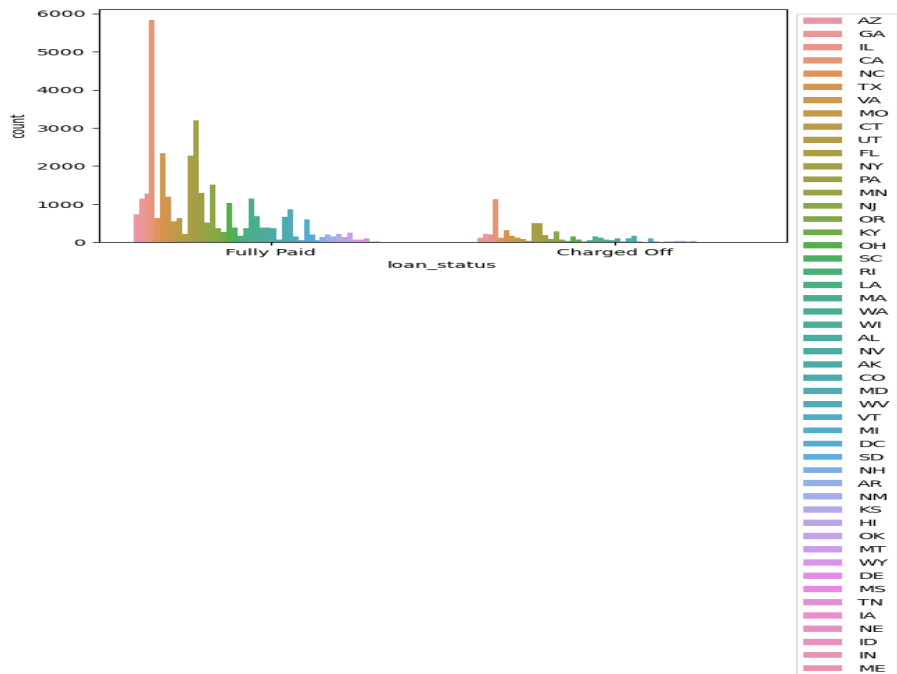
Loan applications who are taking loans for the purpose of debt consolidation is more likely to default the loan



Bivariate Analysis & Conclusions

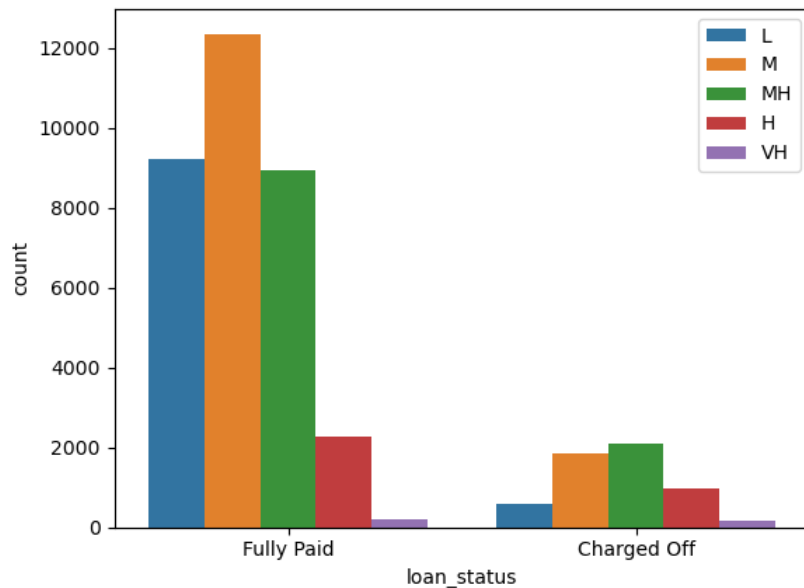
Loan applications from CA addresses are more likely to default the loan

Addr_state vs loan status



Loan applications having an interest rate between M and MH are more likely to default the loan

Interest Rate vs loan status



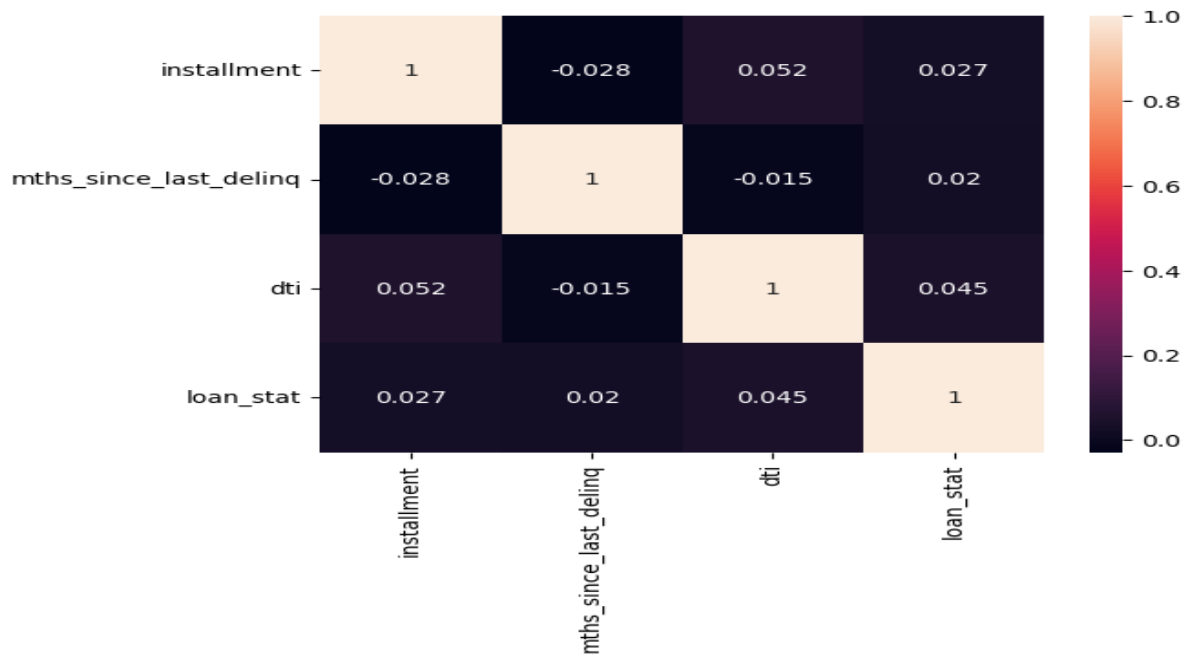
- # L - low - 5 to 9
- # M - Medium- 9 to 13
- # MH - medium high - 13 to 17
- # H - high - 17 to 21
- # VH - very high - 21 to 25

Multivariate Analysis & Conclusions

Multivariate Analysis & Conclusions

With respect to loan_status which is the target - dti has got the highest correlation with loan status.

I.e if the debt to income increases the loan applicant tends to default the loan as he has more debt to be repaid than his income



Thank You!