Lending Club Case Study

Pindi Lova Babu

Lending Club: EDA Case Study

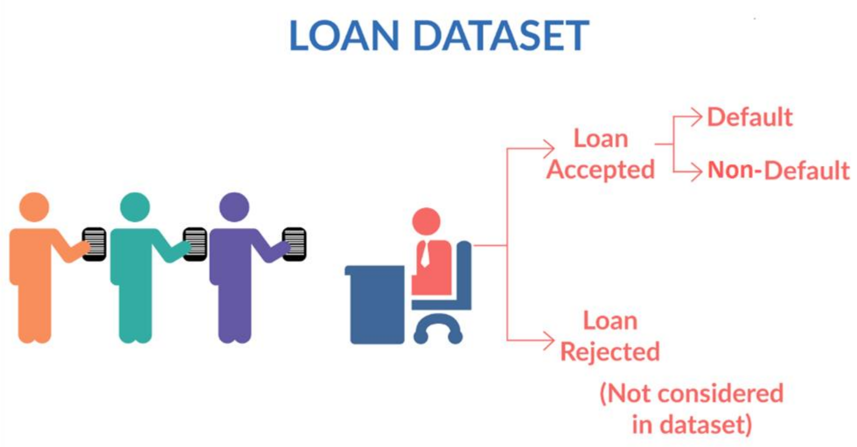
upGrad

What is Lending Club?

Lending Club is a marketplace for personal loans that matches borrowers who are seeking a loan with investors looking to lend money and make a return.

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



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

Main Steps:

1. Understand the data
2. Figure out redundant and invalid columns
3. Clean the dataset which means delete columns which are required.
4. Impute data wherever its required
5. Analyze each column with univariant analysis
6. Bivariant and multi variant analysis
7. Recommendations to lending club

Few Important Variables

Important variables

Loan_amt
Term
Interest_rate
Grade
SubGrade
Annual Income
Purpose of loan
DTI (Debt to Income)
Emp_Lenght
Loan_Date(Month)
Home_Ownership
Verification_Status

Term:
- 36 Months
- 60 Months

Grade:
- A
- B
- .
- .

SubGrade:
- 1
- 2
- .
- .

Verification_Status:
- Income Verified (Pay Slips, Tax Returns)
- Income Source Verified (3rd Party vendors has checked)
- Not Verified

Customer behaviour variables

delinq_2yrs
earliest_cr_line
inq_last_6mths
open_acc
pub_rec
revol_bal
revol_util
total_acc
out_prncp
out_prncp_inv
total_pymnt
total_pymnt_inv
total_rec_prncp
total_rec_int
total_rec_late_fee
recoveries
collection_recovery_fee
last_pymnt_d
last_pymnt_amnt
last_credit_pull_d
application_type

the customer behavior variables are not available at the time of loan application, and thus they cannot be used as predictors for credit approval.

variables such as acc_now_delinquent, chargeoff within 12 months etc. (which are related to the applicant's past loans) are available from the credit bureau.

Data Analysis: Univariate Analysis

- For univariate analysis, you may check the default rate across various categorical features.
- For continuous features, you may perform binning and then you may perform univariate analysis.

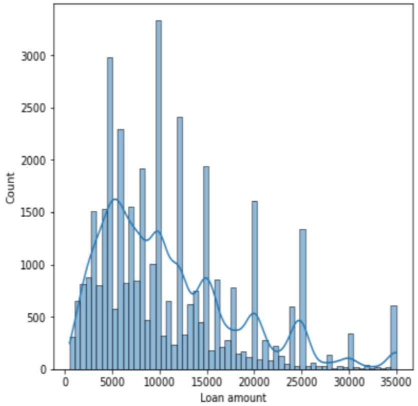
Data Analysis: Bivariate Analysis

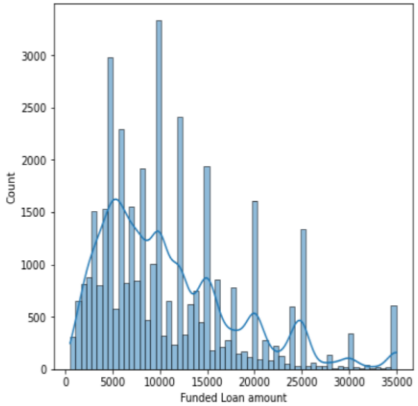
- Here you may choose two features to understand the Default variable

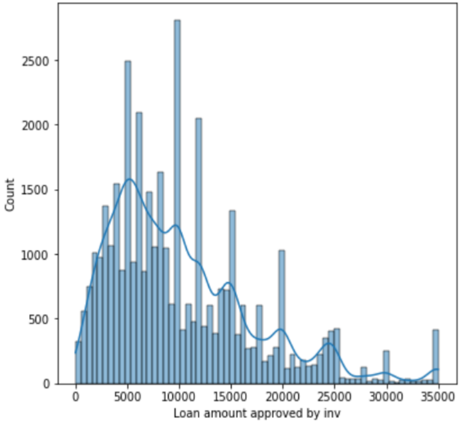
Loan Amount vs Funded amount

Observations:

Loan Amount applied, funded amount and Funded amount investment has the almost same distribution which means that most of the borrowers got approved for applied loans amounts







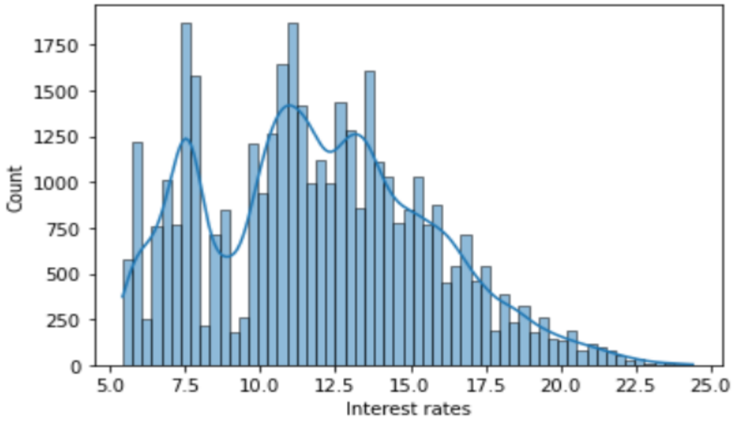
Rate of interest and DTI

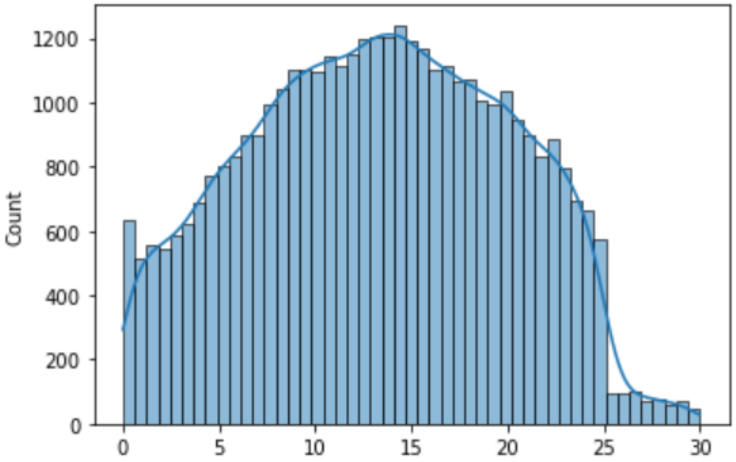
upGrad

Observations:

- Most of the loans taken for interest rate of 7.5% to 17.5%

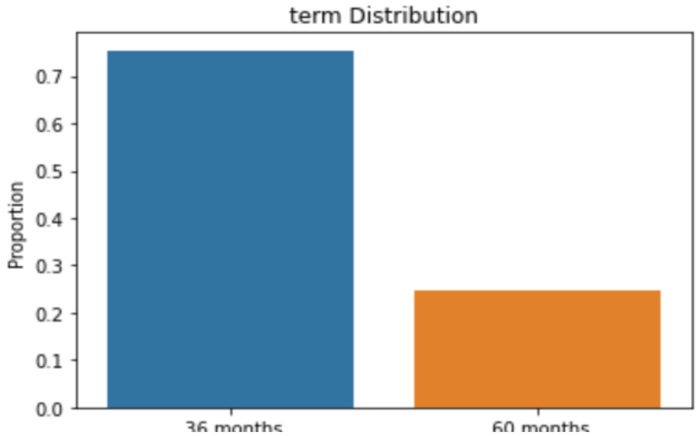
- Looks like there are no outliers and the distribution is very much similar to normal distribution. This is good sign that all the loans are given to barrower's who have Debt to Income ration less than 30.

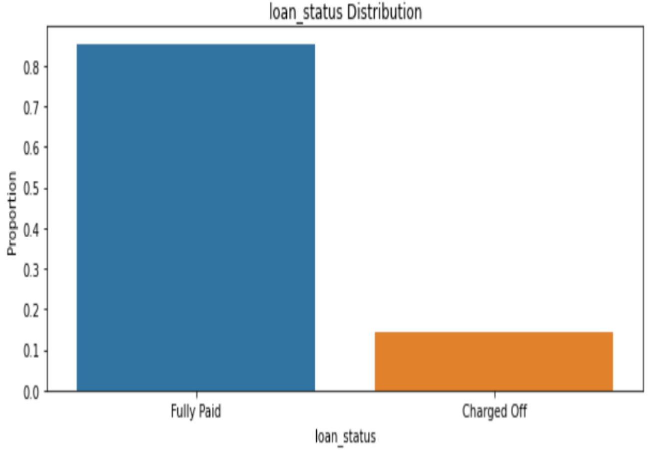




Term distribution

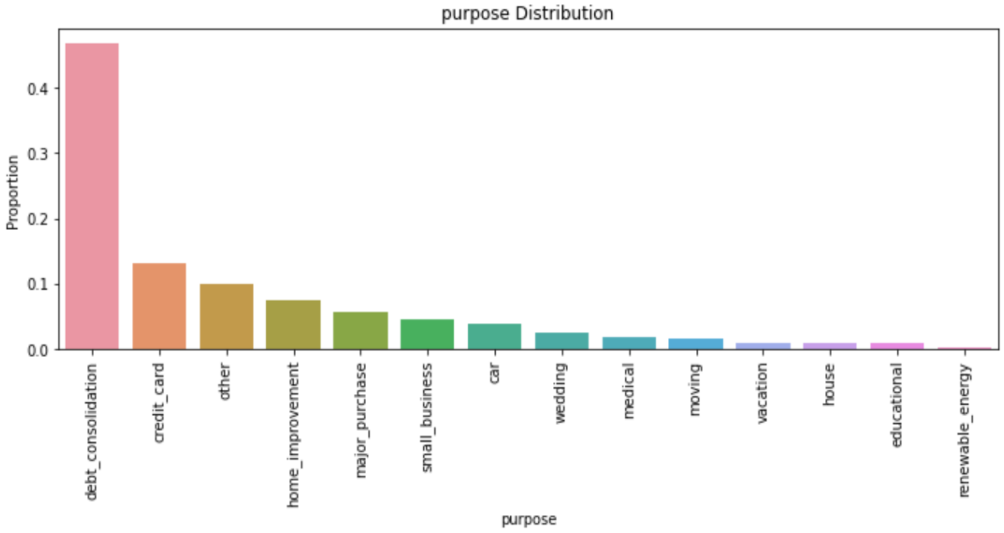
This plot clearly say that more loans sanctioned for term of 36 months



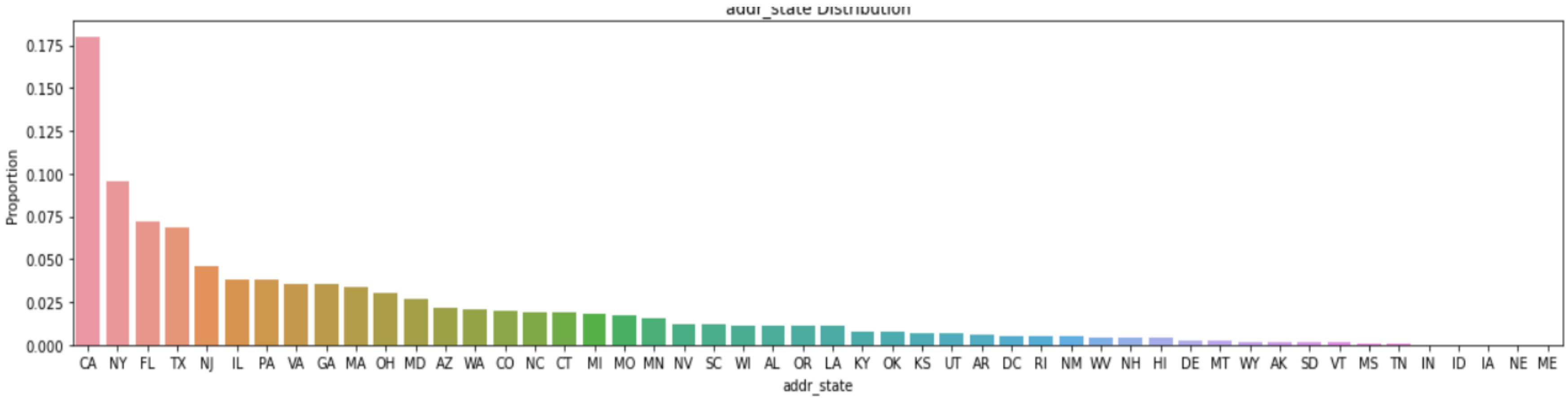


Observation:

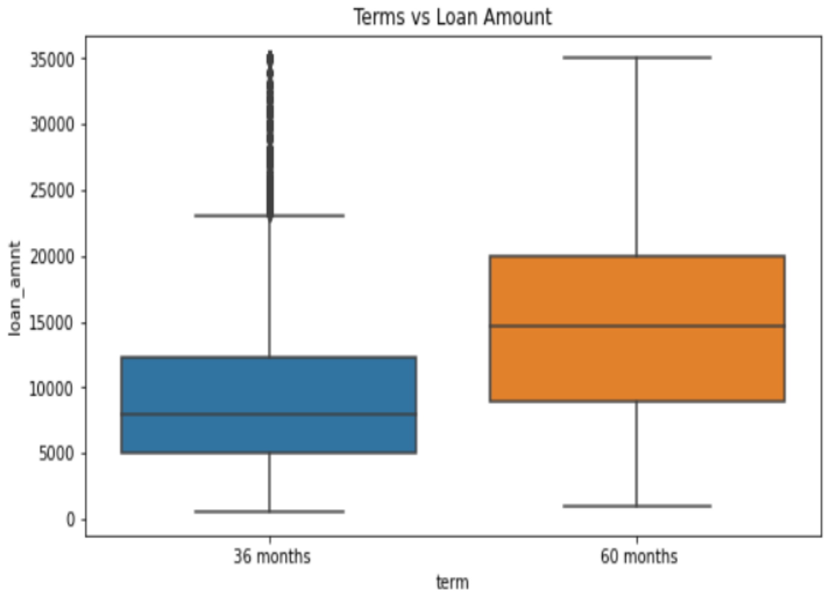
More loans taken as purpose of debt consolidation

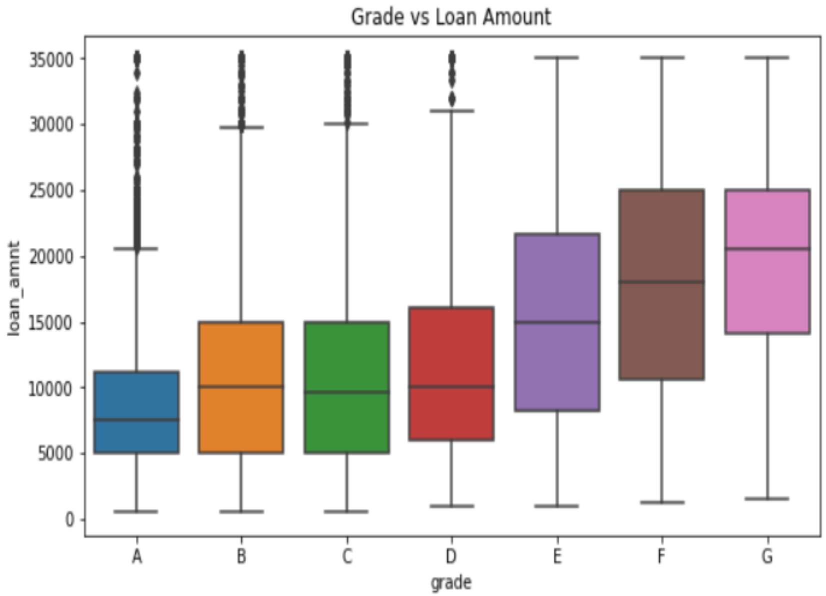


Addr_state

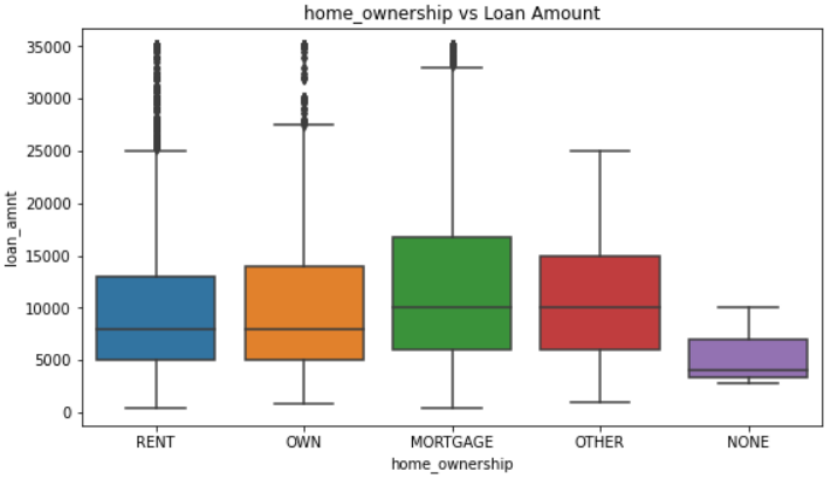


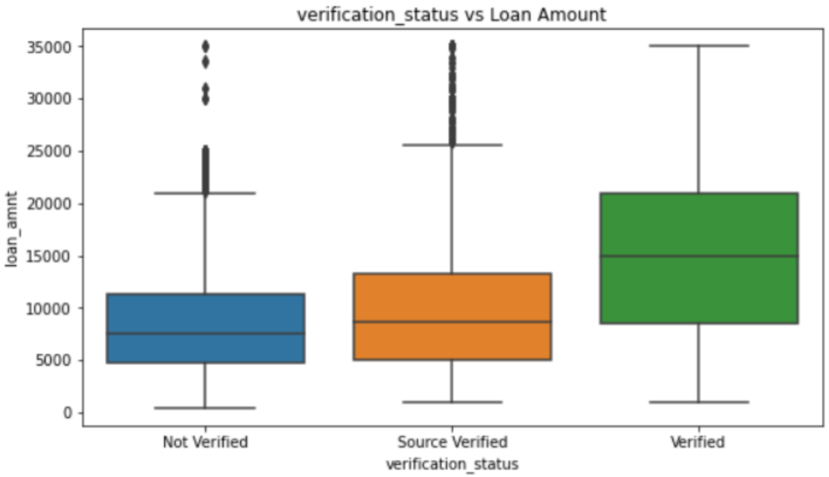
Observation: People under state CA are taken more loans



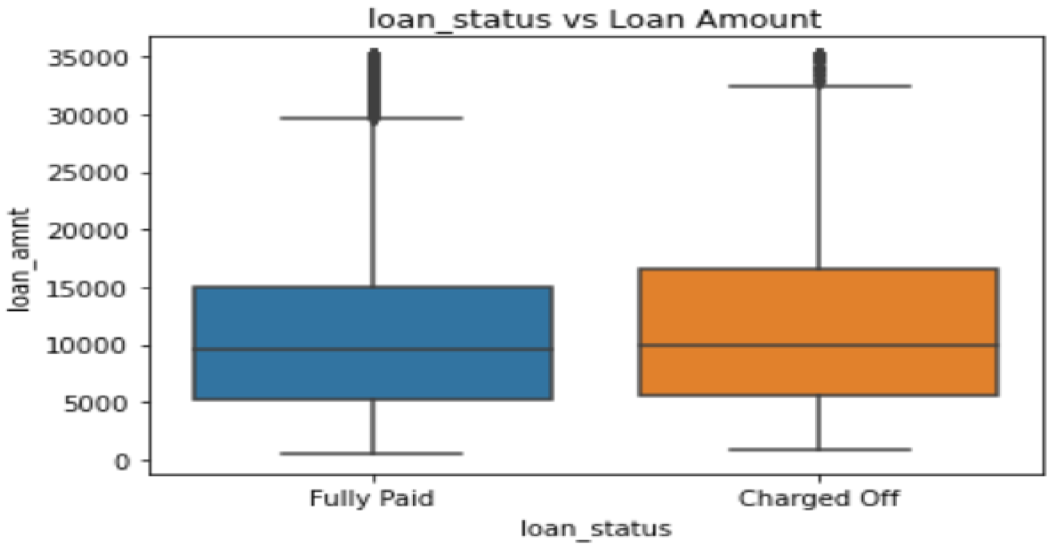


Above graphs clearly shows that large amounts given to 60 months terms, similarly large amounts given to low grades

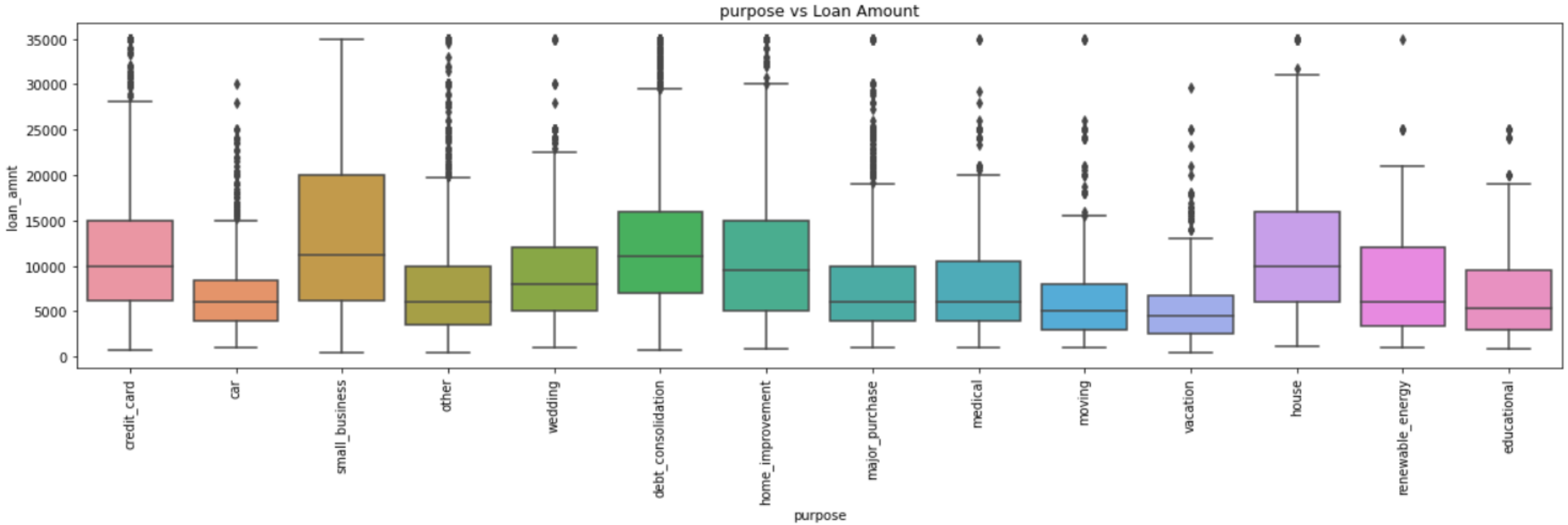




Above graphs clearly says that larger amounts given on Mortage and own house, Rented. similarly, who all are verified successfully lending club has givn larger amounts

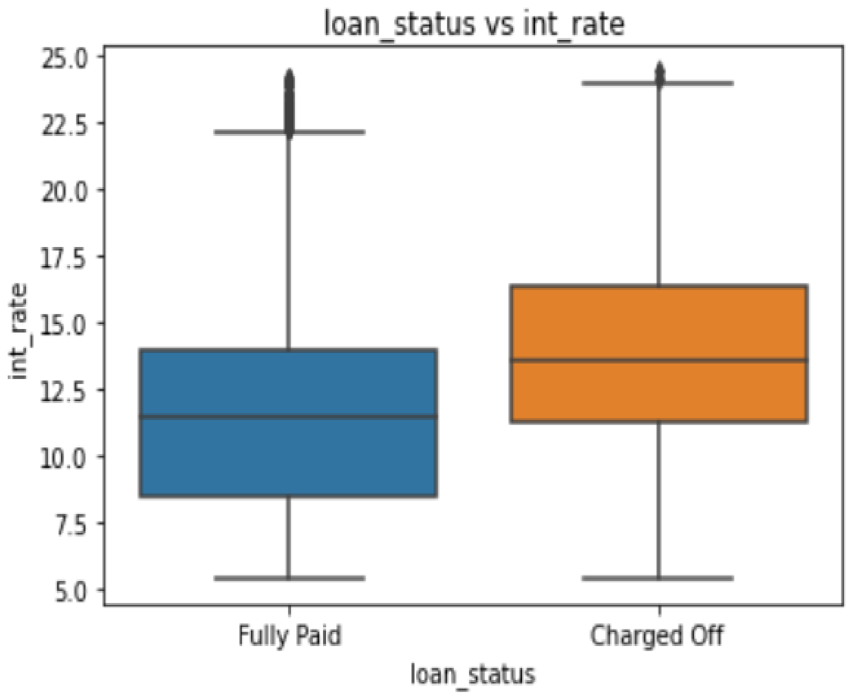


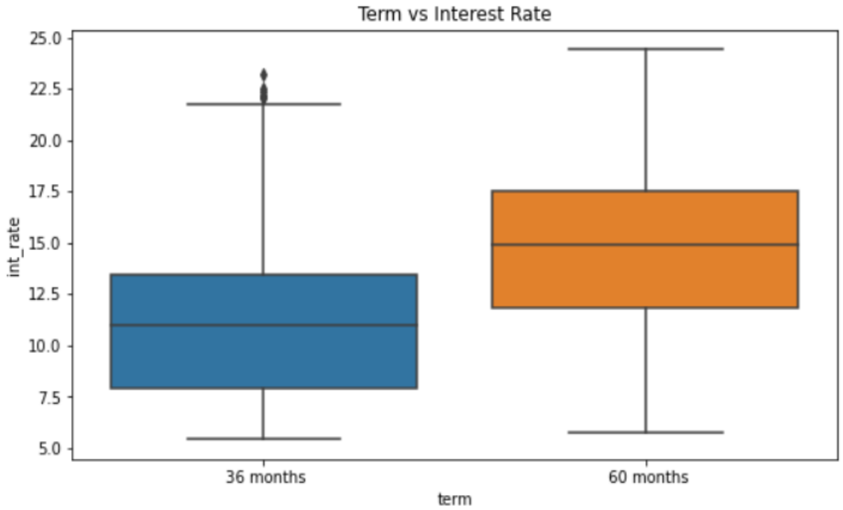
This quite interesting that loans with charged off are given with larger amounts

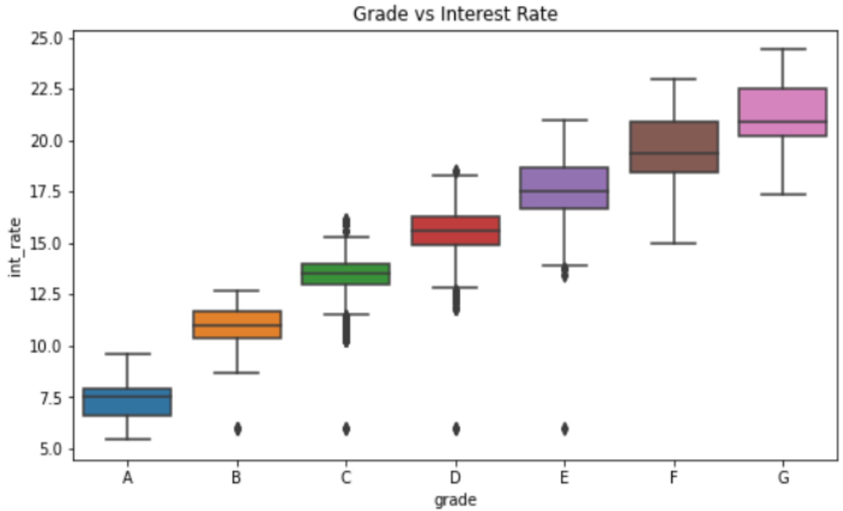


Larger amounts of loans given on purpose of small business, house and debt consolidation purposes

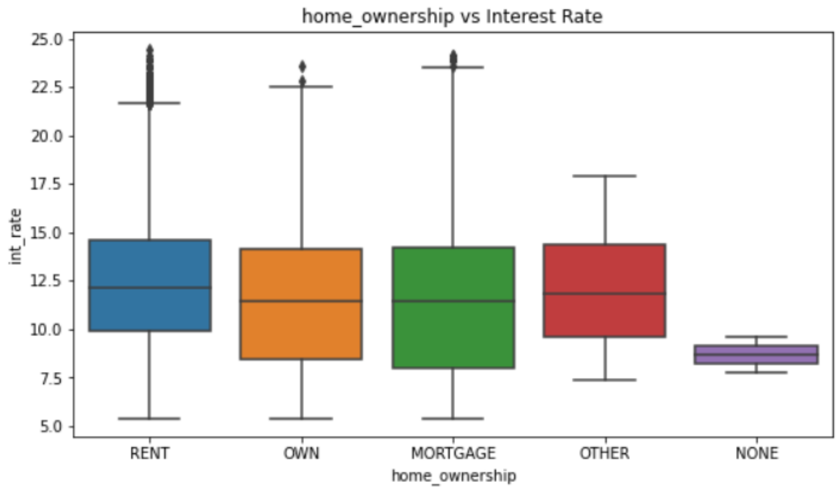
This is more useful information; this says that who are taken loans with more interest rates they likely to default

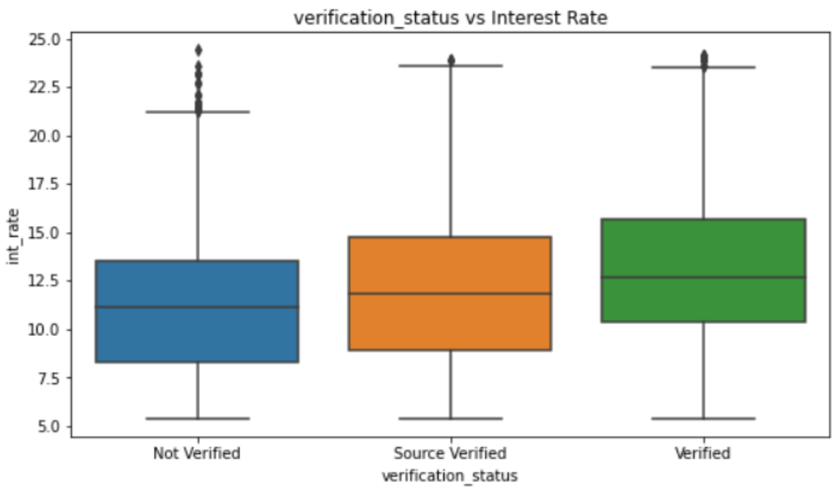






This is useful information that more interest charged on long terms loans, At the same time the loans given with more interest rates are likely to defaulted

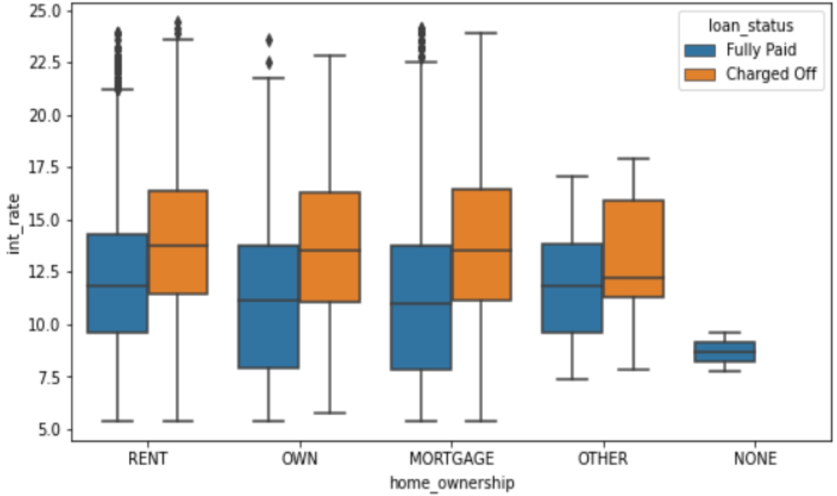


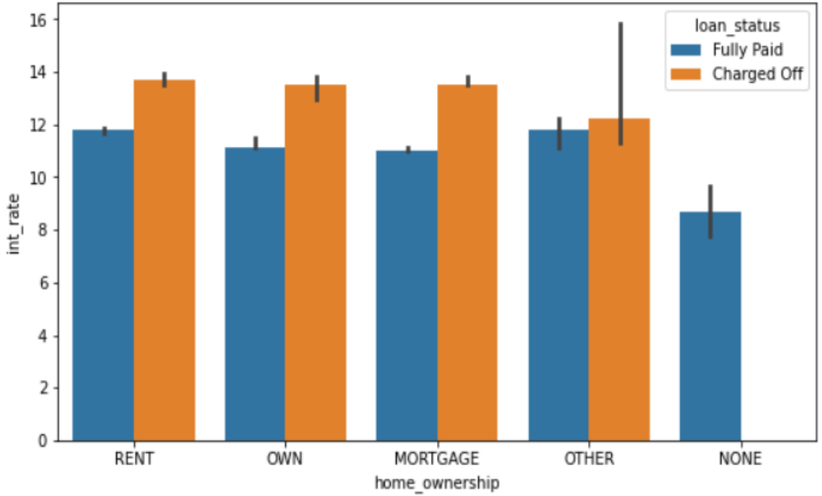


1. Loans are given with more interest rate for who are with Rented homes, this can again lead to default loans since all we know loans sanctioned with more interest rates are defaulted
2. one more important point here is that applicants who all done with verifications are got charged with more interest. And Not verified applicants got less interest rates

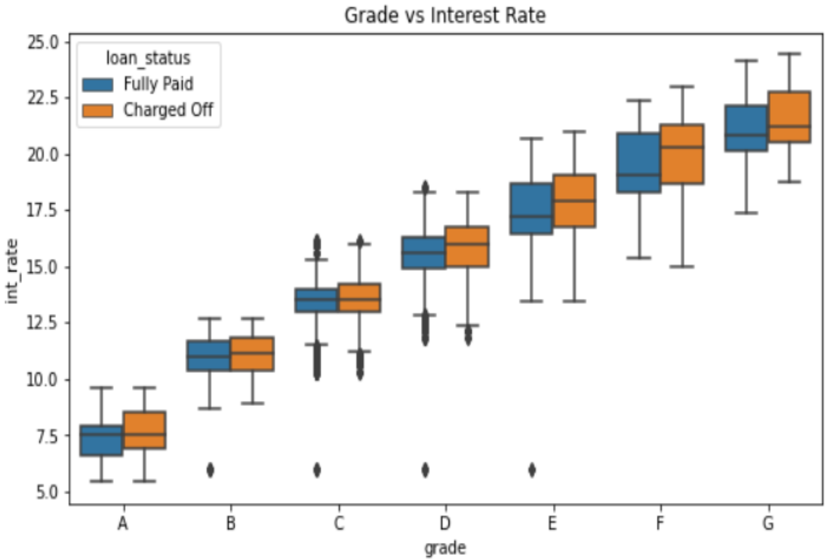
upGrad

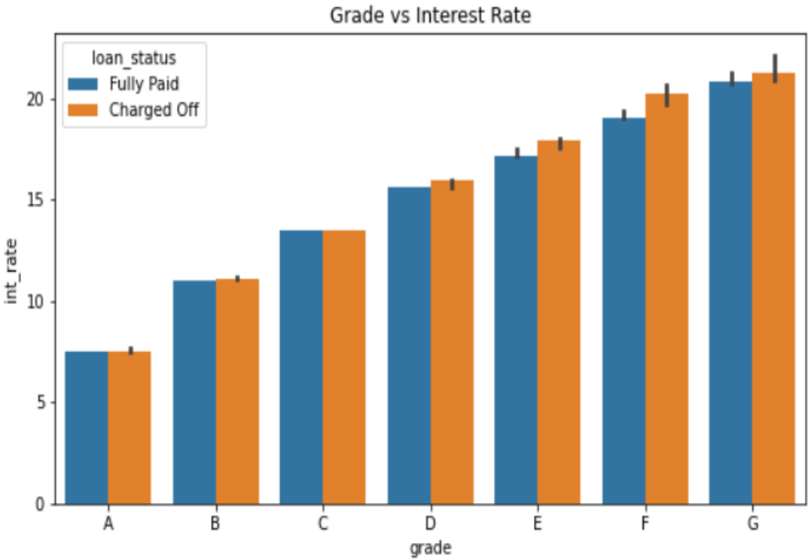
home_ownership vs Interest Rate



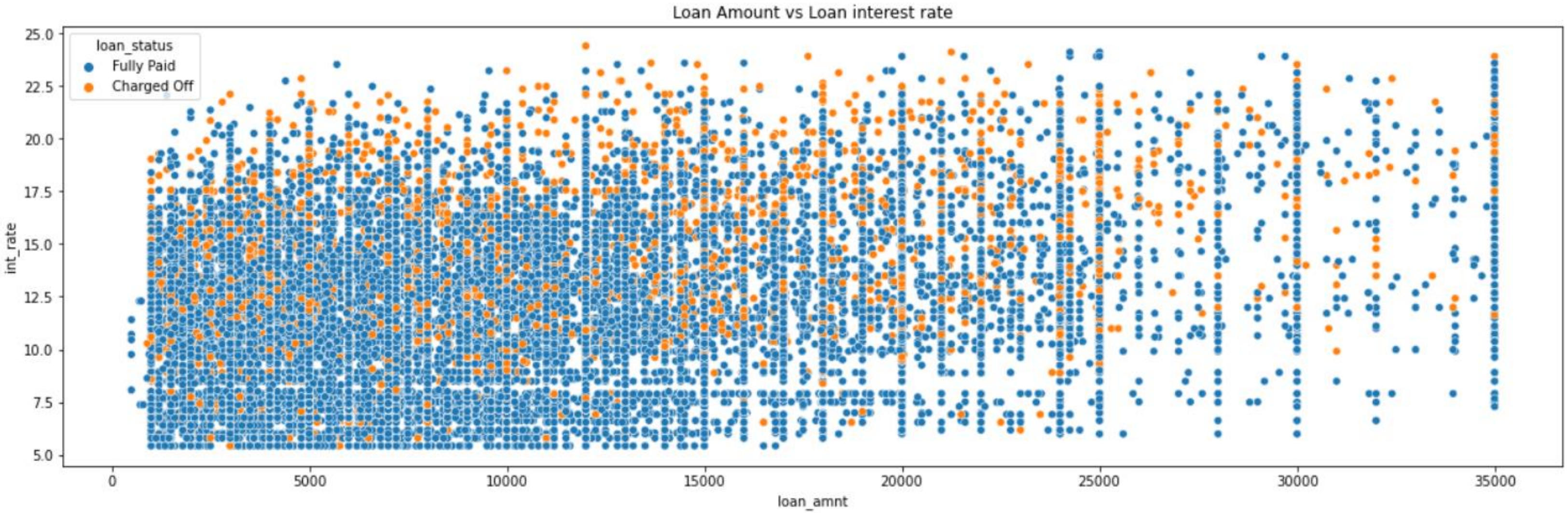


Irrespective home ownership, its clearly says that of high ROI will lead default the loan





Its clearly shows as per the category, defaults loans can be identified, Low grade like G have possibilities to default the loan



This scatter plot shows that there more loans defaulted when ROI is more

Recommendations to Lending club

upGrad

1. if Debt to income ratio is more, there are more possibilities to default the loan
2. Proper scrutiny is required when loans are giving for high rate of interest, since observed more defaulters when ROI is high
3. Loan taken for purpose of credit card consolation or debt consolidation having more defaults
4. Observed more defaults under rented house applicants compared to own house.
5. Loan taken as larger amount with 60 terms having more defaults
6. Low grade having more defaults, its good that lending club taking consideration on this.

upGrad

#LifeKoKaroLift

Thank You!