Lending Club Case Study

1 Data Understanding

- loan.csv has 39717 rows x 111 columns
- Here the main concern is to identify the loan application who is likely to “Charge-off”

2 Data Cleaning

- Out of 111 columns , 57 columns holds only null values

- They are ['mths_since_last_major_derog', 'annual_inc_joint', 'dti_joint', 'verification_status_joint', 'tot_coll_amt', 'tot_cur_bal', 'open_acc_6m', 'open_il_6m', 'open_il_12m', 'open_il_24m', 'mths_since_rcnt_il', 'total_bal_il', 'il_util', 'open_rv_12m', 'open_rv_24m', 'max_bal_bc', 'all_util', 'total_rev_hi_lim', 'inq_fi', 'total_cu_tl', 'inq_last_12m', 'acc_open_past_24mths', 'avg_cur_bal', 'bc_open_to_buy', 'bc_util', 'mo_sin_old_il_acct', 'mo_sin_old_rev_tl_op', 'mo_sin_rcnt_rev_tl_op', 'mo_sin_rcnt_tl', 'mort_acc', 'mths_since_recent_bc', 'mths_since_recent_bc_dlq', 'mths_since_recent_inq', 'mths_since_recent_revol_delinq', 'num_accts_ever_120_pd', 'num_actv_bc_tl', 'num_actv_rev_tl', 'num_bc_sats', 'num_bc_tl', 'num_il_tl', 'num_op_rev_tl', 'num_rev_accts', 'num_rev_tl_bal_gt_0', 'num_sats', 'num_tl_120dpd_2m', 'num_tl_30dpd', 'num_tl_90g_dpd_24m', 'num_tl_op_past_12m', 'pct_tl_nvr_dlq', 'percent_bc_gt_75', 'tot_hi_cred_lim', 'total_bal_ex_mort', 'total_bc_limit', 'total_il_high_credit_limit']

continued..

- Here our main Analysis is for new loan applicants. Hence all the row with the "*loan_status*" as "*Current*" can be removed. (Now the shape of dataset is (38577, 57))

- mapping the categorical values column "loan_status" to integer for easy analysis and visualization
  - Fully Paid ==> 0
  - Charged Off ==> 1

- Removing the columns which has single unique value has doesn't add much value to the Analysis

  Columns - *pymnt_plan,application_type,initial_list_status,policy_code*

continued..

- **the customer behavior variables are not available at the time of loan application, and thus they cannot be used as predictors for credit approval.**
- removing the following customer behavior variables
- delinq_2yrs - The number of 30+ days past-due incidences of delinquency in the borrower's credit file for the past 2 years
- earliest_cr_line - The month the borrower's earliest reported credit line was opened
- inq_last_6mths - The number of inquiries in past 6 months (excluding auto and mortgage inquiries)
- open_acc - The number of open credit lines in the borrower's credit file.
- pub_rec - Number of derogatory public records
- revol_bal - Total credit revolving balance

continued..

- revol_util - Revolving line utilization rate, or the amount of credit the borrower is using relative to all available revolving credit.
- total_acc - The total number of credit lines currently in the borrower's credit file
- out_prncp - Remaining outstanding principal for total amount funded
- out_prncp_inv - Remaining outstanding principal for portion of total amount funded by investors
- total_pymnt - Payments received to date for total amount funded
- total_pymnt_inv - Payments received to date for portion of total amount funded by investors
- total_rec_prncp - Principal received to date
- total_rec_int - Interest received to date

continued..

- total_rec_late_fee - Late fees received to date
- recoveries - post charge off gross recovery
- collection_recovery_fee - post charge off collection fee
- last_pymnt_d - Last month payment was received
- last_pymnt_amnt - Last total payment amount received
- last_credit_pull_d - The most recent month LC pulled credit for

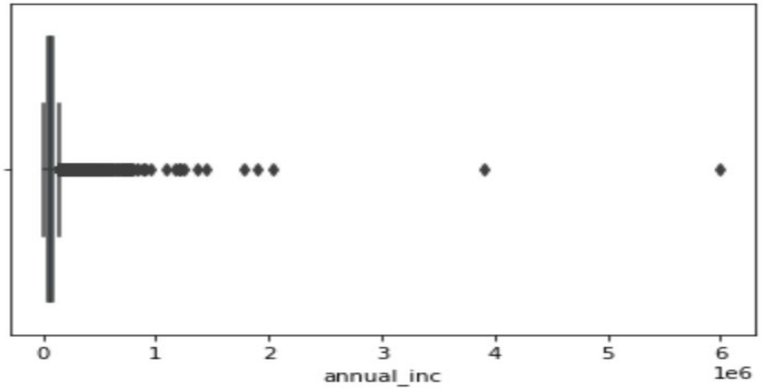
- **Now the shape of dataset is (38577, 32)**

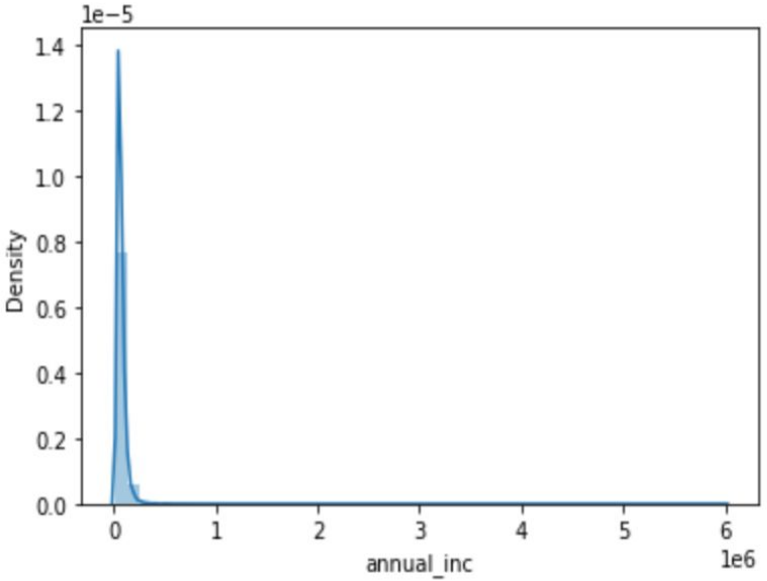
2.1 converting Object type to appropriate Data type for better Analysis

- In column “term” the string is converted to integer by removing “ months”
- column “int_rate” the string is converted to float by removing “%”
- column “Annual_inc” is converted to integer
- “Issue_d” is converted to datetime object
- Removing the columns which has single unique value and ‘nan’ -
  - Acc_now_delinq,
  - delinq_amnt
  - Collections_12_mths_ex_med
  - Chargeoff_within_12_mths
  - tax_liens'

3 Univariate Analysis

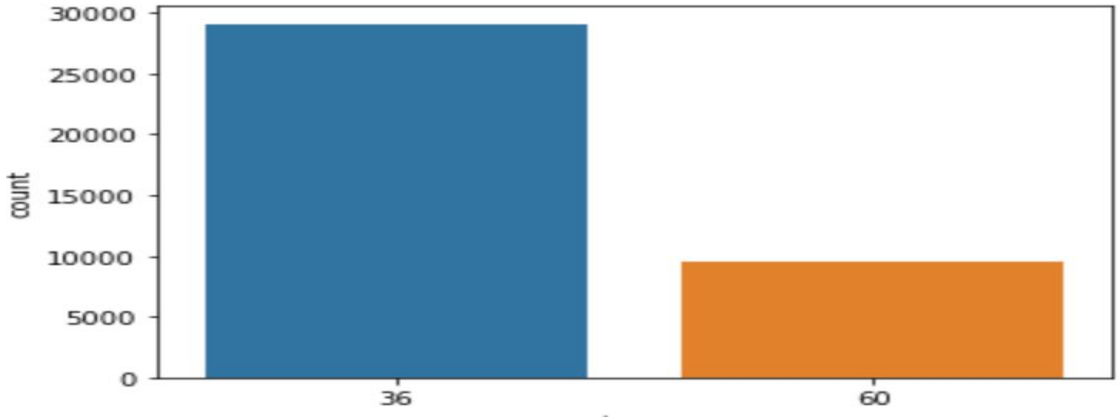
3.1 annual_inc





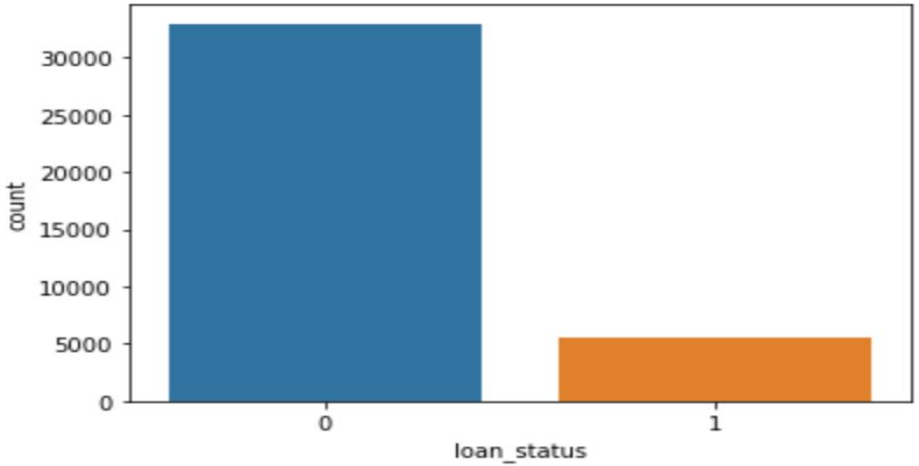
The above plots show that "annual_inc" has lot of mean values for analysis

3.2 term



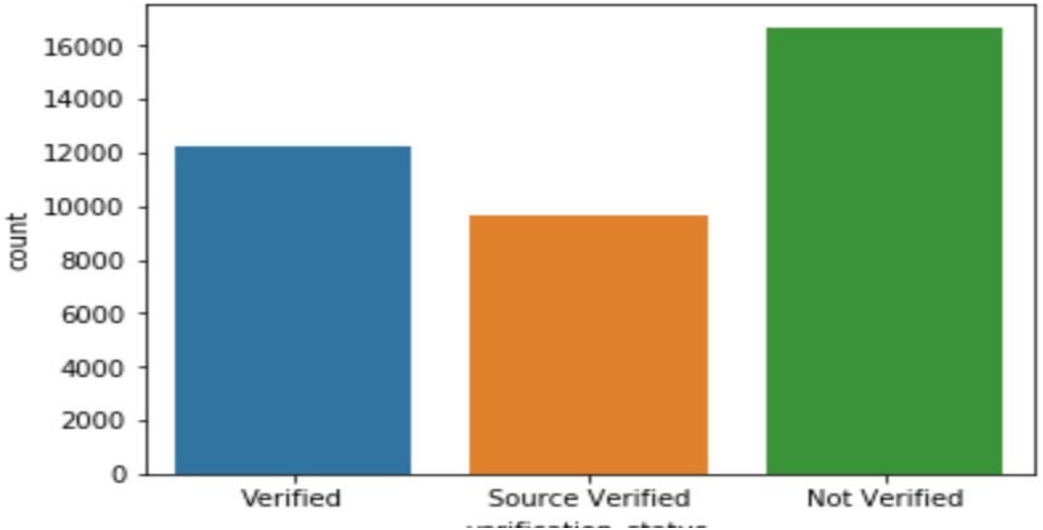
the above plot shows that the dataset mostly contains the records for the loan opted for 36 months

3.3 loan_status



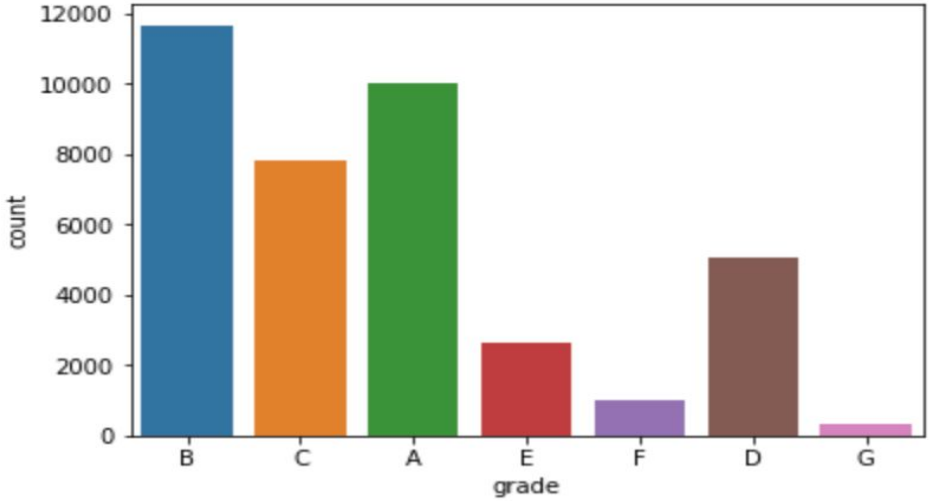
the above plot shows that the dataset mostly contains the records for the fully paid categories

3.4 verification_status



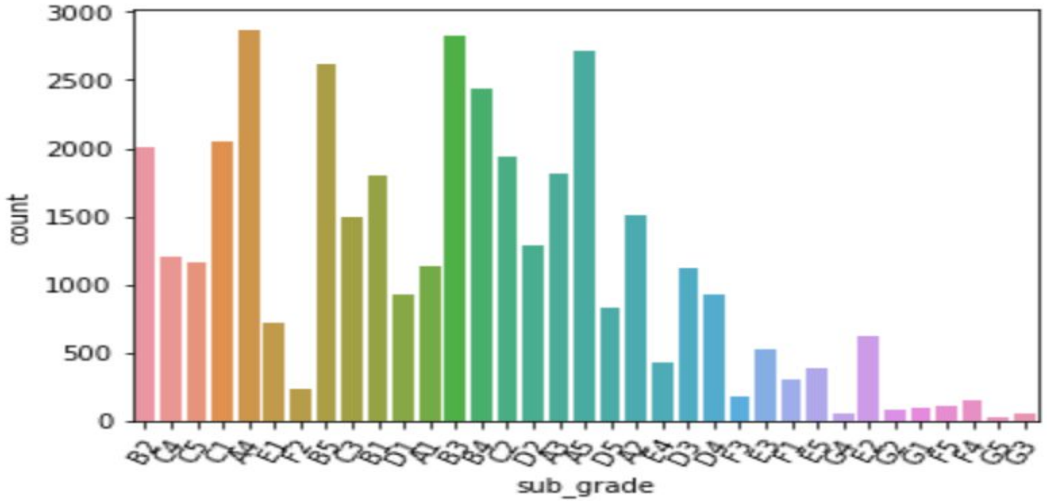
the above plot shows that the dataset mostly contains the records for the “Not Verified” categories

3.5 grade



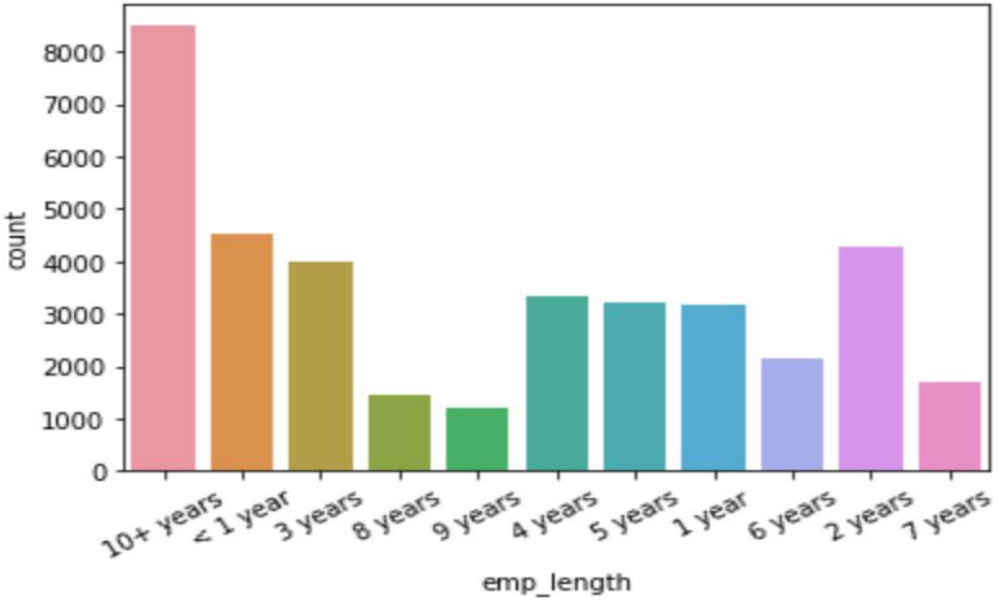
the above plots shows that the dataset has the most occurences of Grade B

3.6 sub_grade



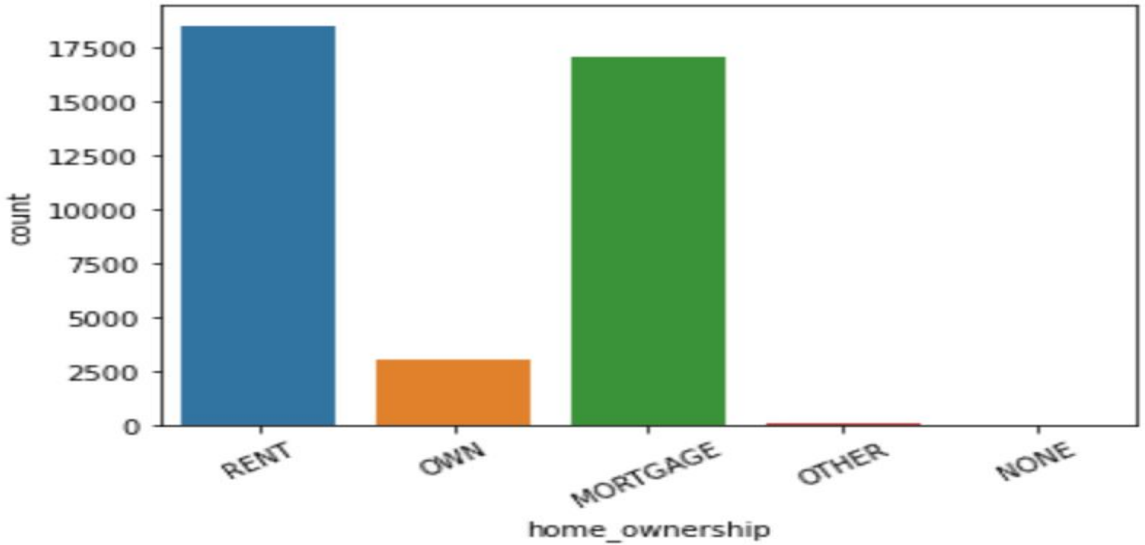
the above plots shows that the dataset has the most occurences of subgrade is A4

3.7 emp_length



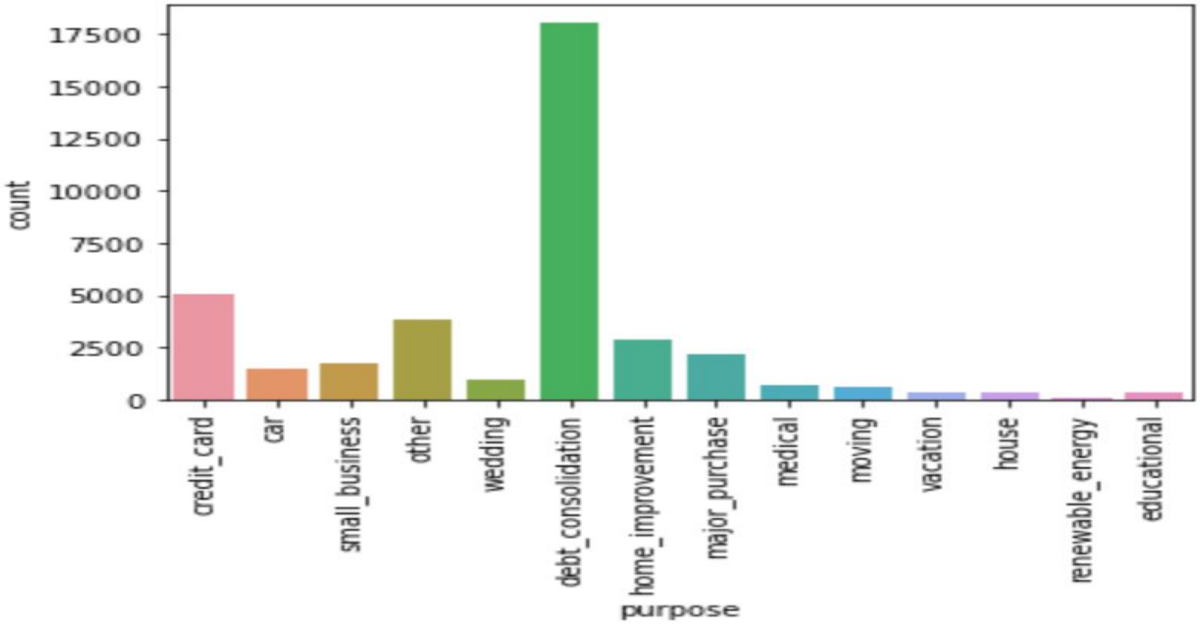
Most of the borrowers who applied loans have 10+ years of experience

3.8 home_ownership



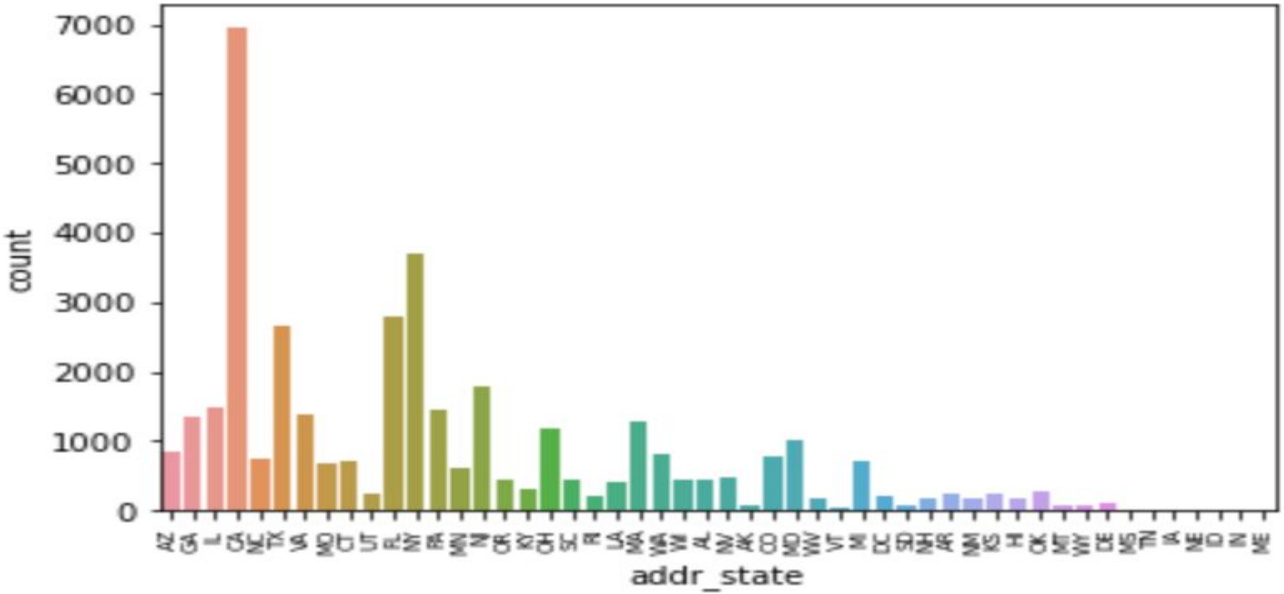
Most of the borrowers who resides in rented property

3.9 purpose



Most of the borrowers applied loan for debt consolidation

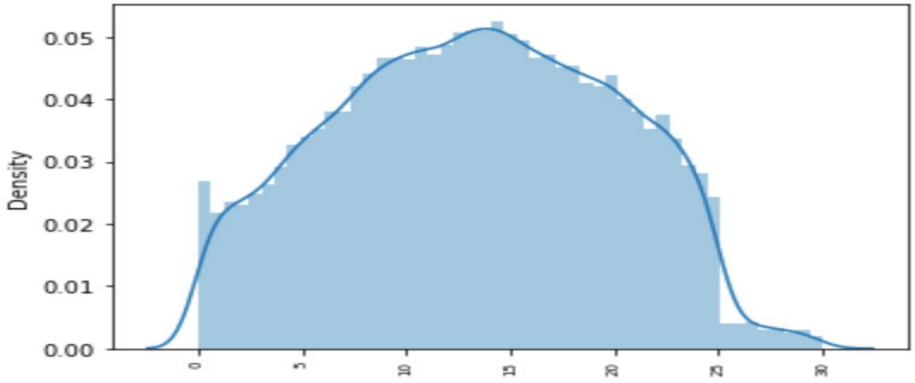
3.10 addr_state



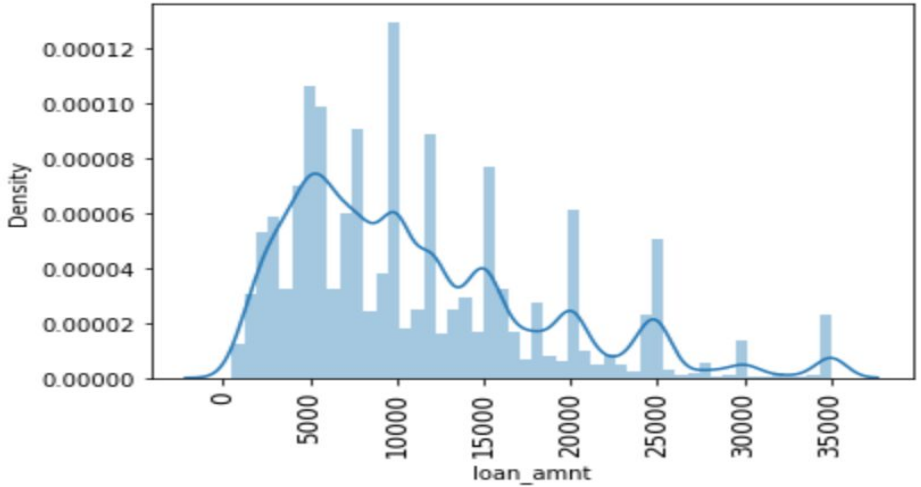
Most of the borrowers applied loan from CA

3.11 distplot

Dti

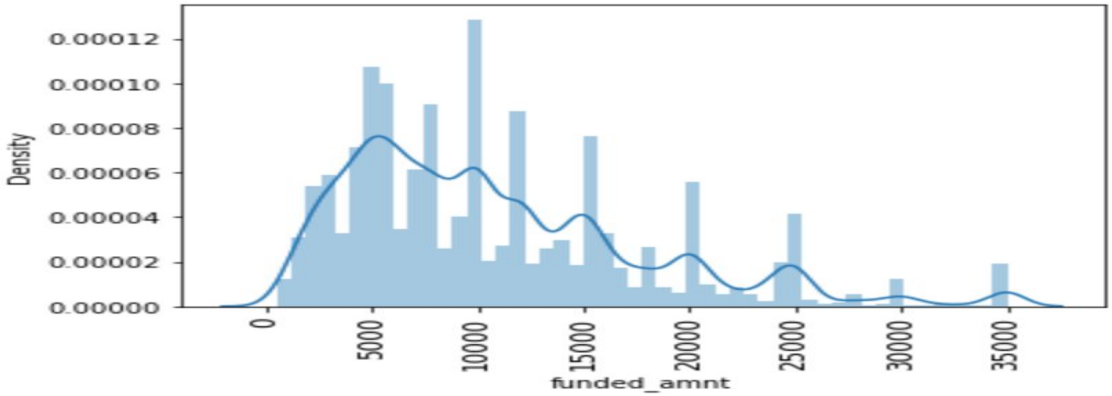


Loan_amnt

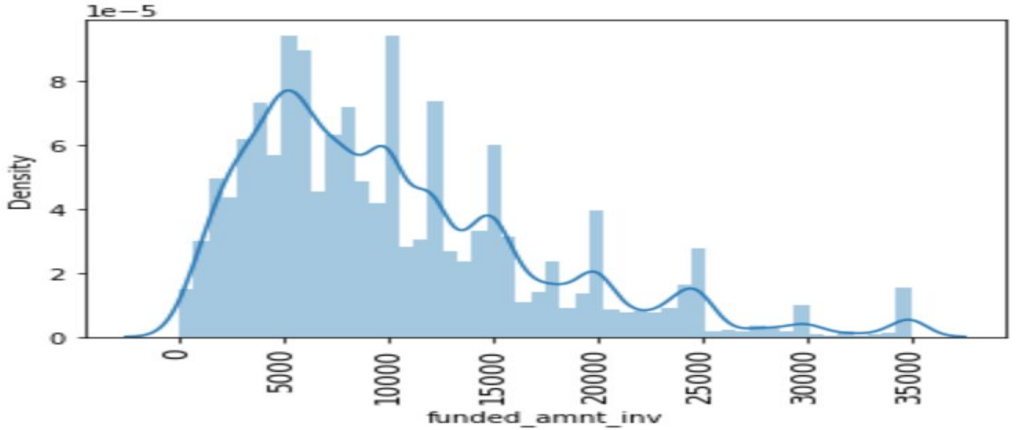


continued..

Funded_amnt

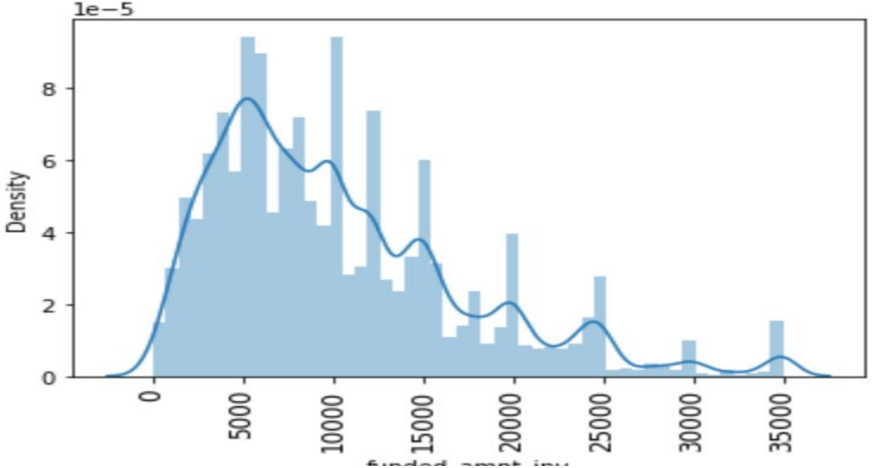


Funded_amnt_inv

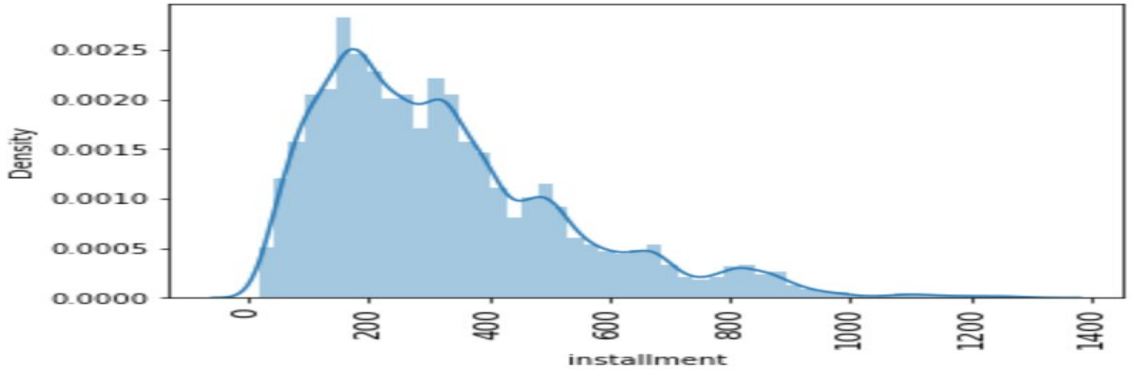


continued..

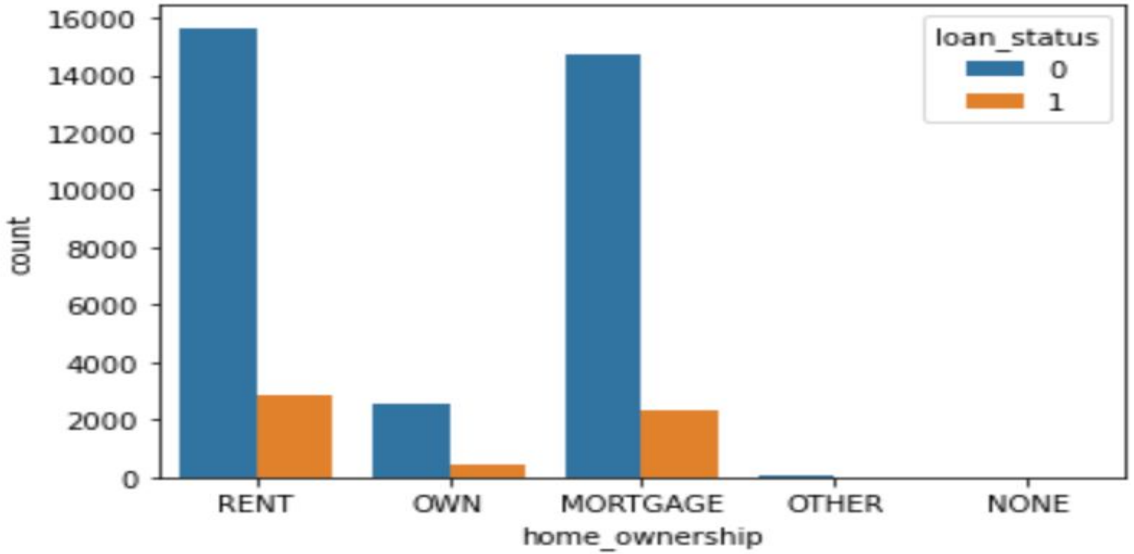
int_rate



Installment

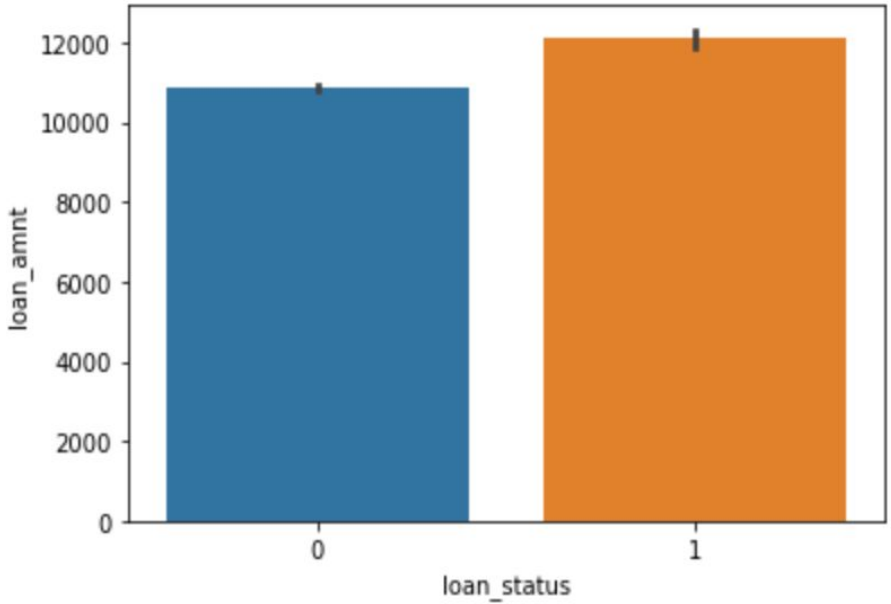


4 bivariate/ Multivariate Analysis



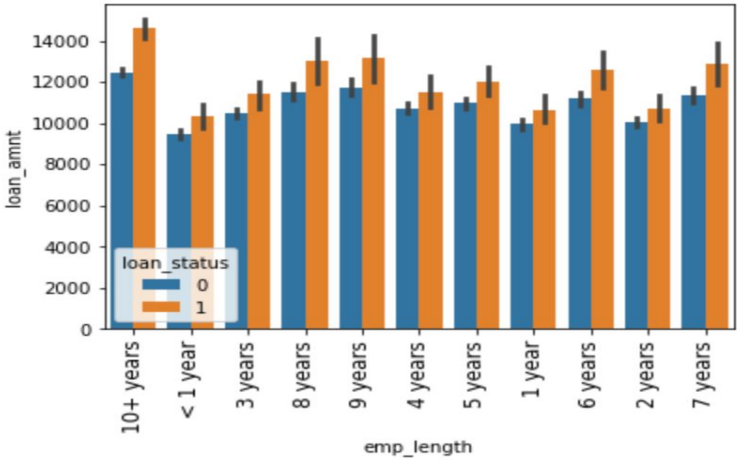
Analysis: The borrowers who has own house has the less changes of defaulting the loan

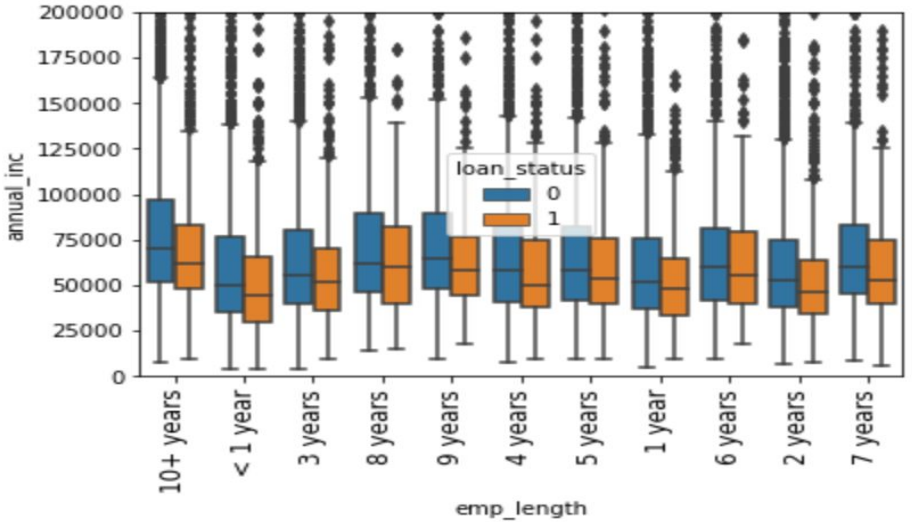
Recommendation: The borrowers with own house can be granted loan



Analysis: The chances of defaulting is high when loan amount is greater than 11000

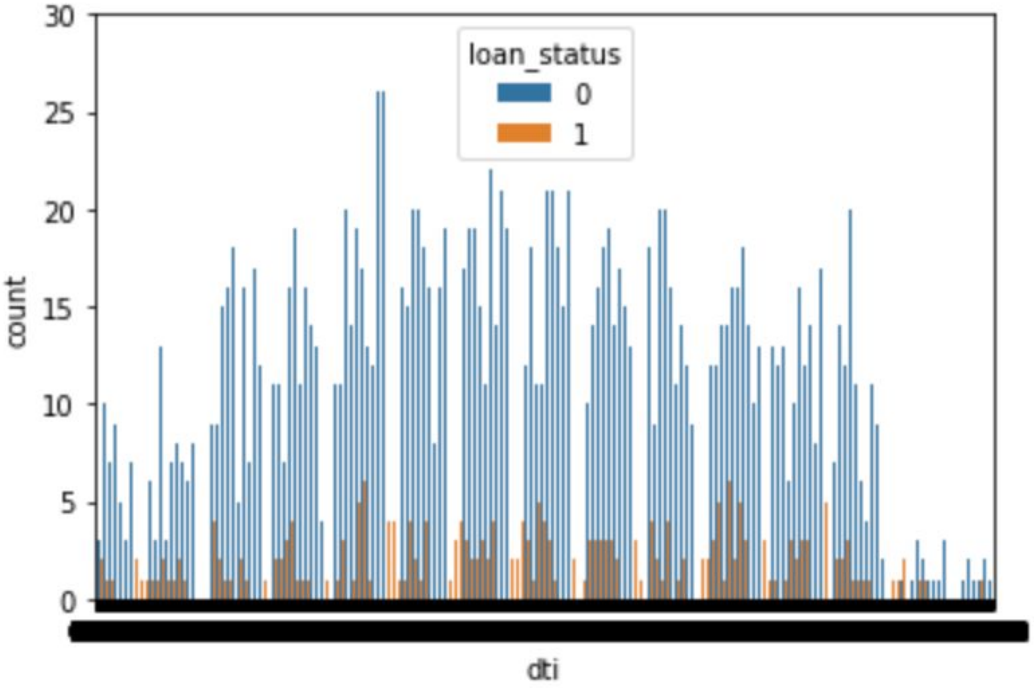
Recommendation: it is always recommended to consider multiple parameters when loan is offered above 11000





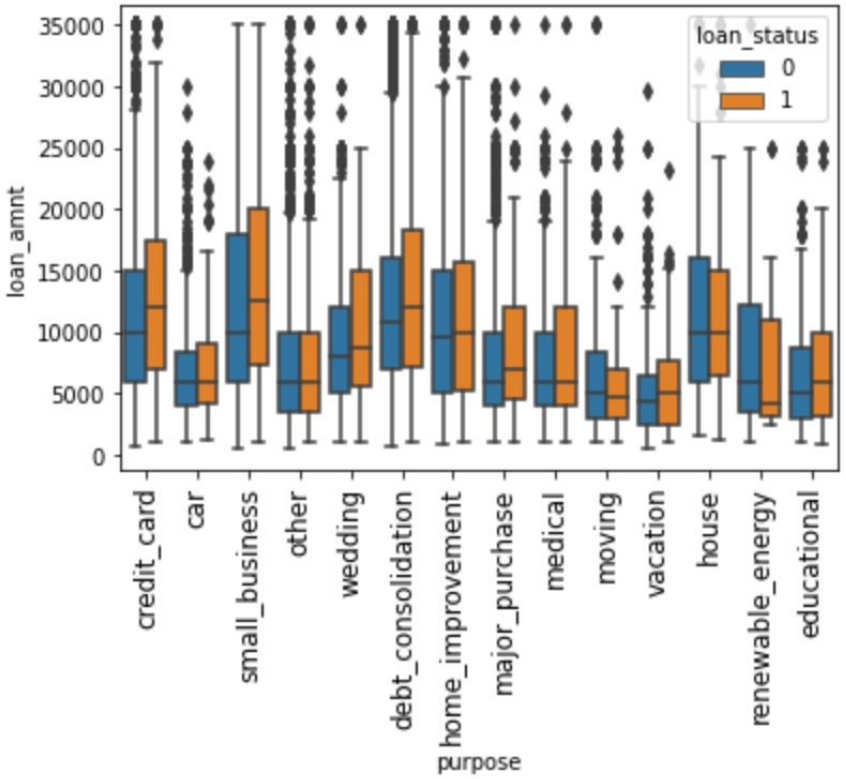
Analysis:the borrowers with 10+ years experience are offered higher loans

Recommendation: The borrowers with annual income greater than 90000 income can be approved loans as they are not defaulting



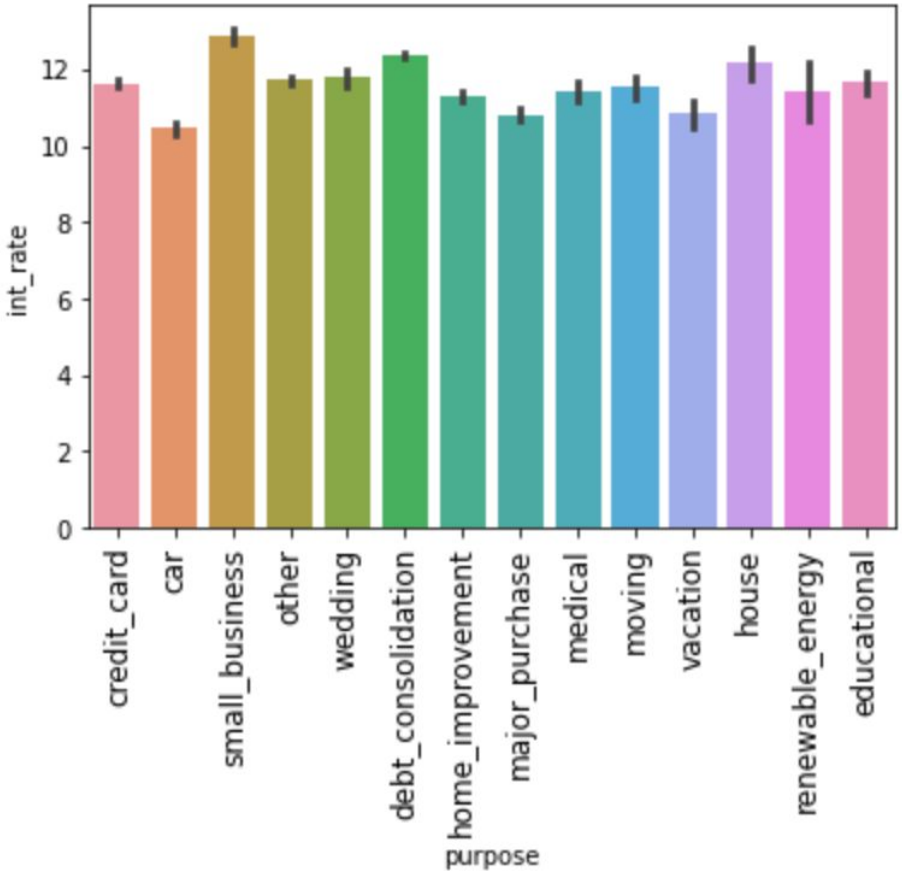
Analysis:dti is directly proportional to loan status

Recommendation: The borrowers with dti greater than 7 can be offered loan as they will not default

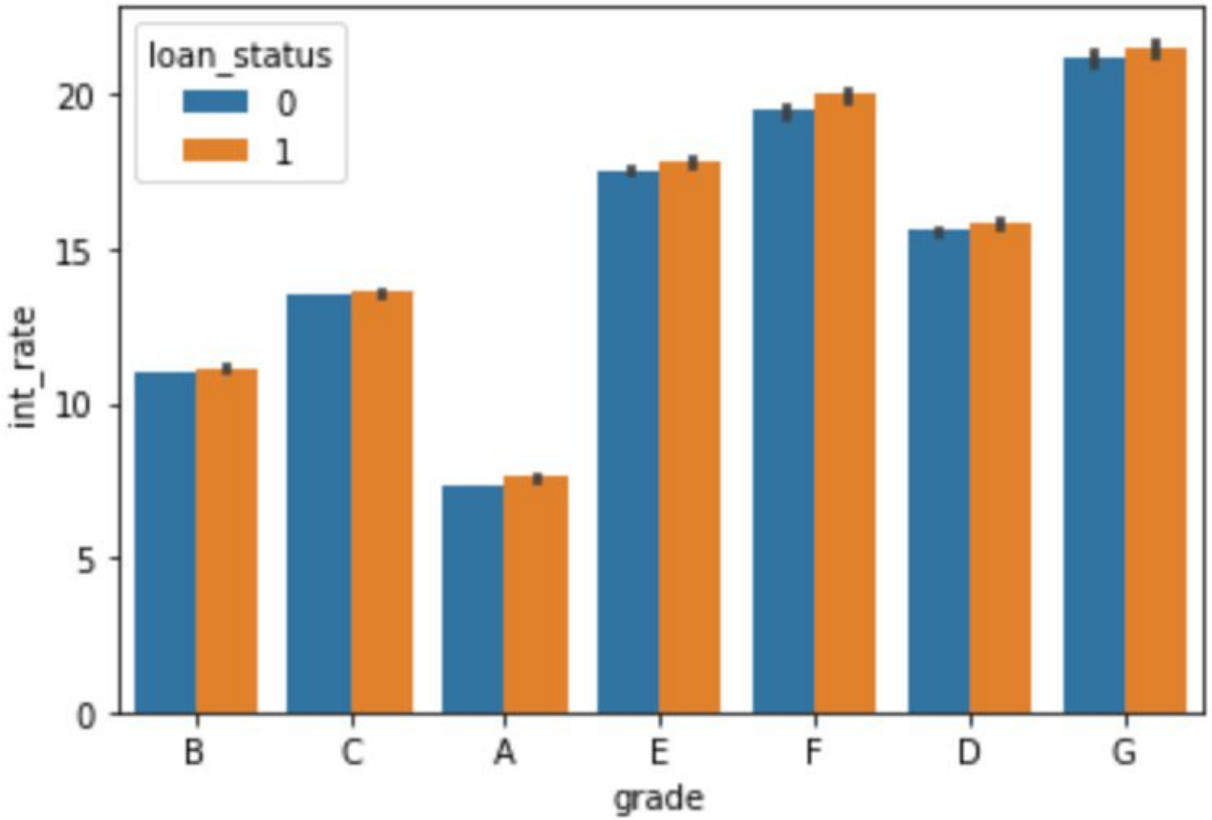


Analysis:the chances of defaulting is less when the purpose is "renewable_enery","house" and "moving"

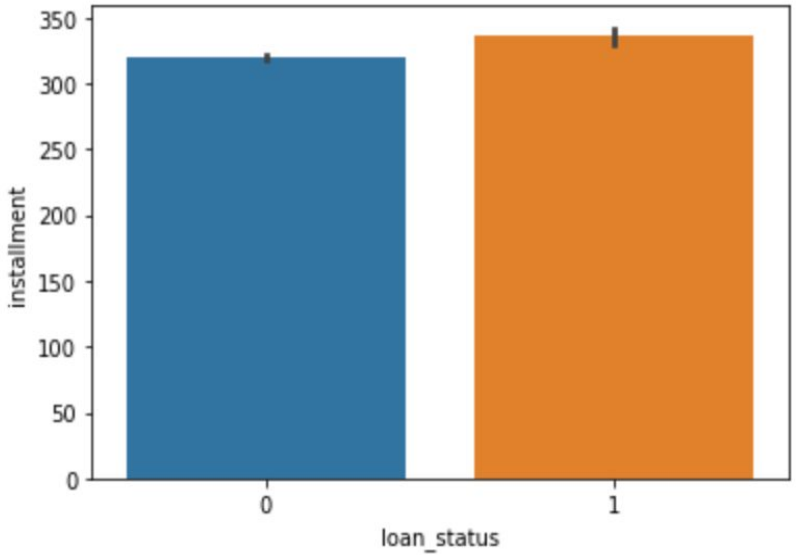
Recommendation:Loans on the purpose "renewable_enery","house" and "moving" are profitable as they are less likely to default

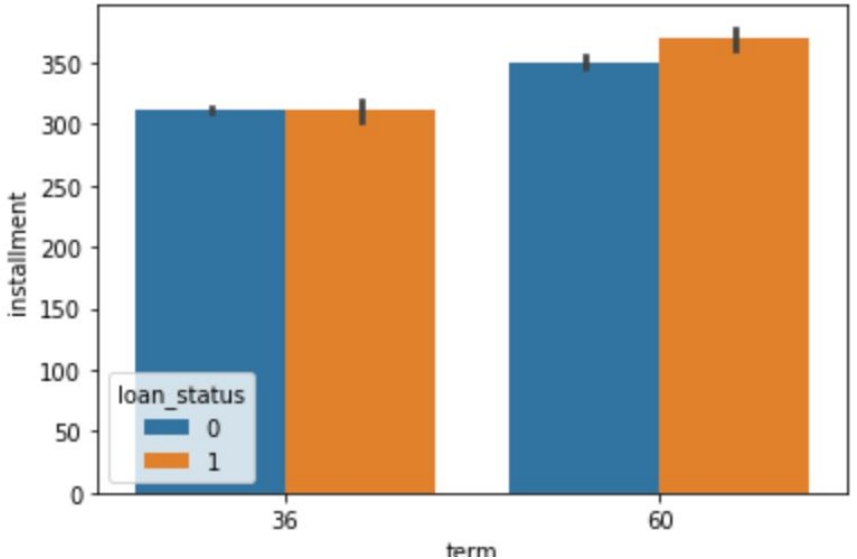


Analysis:The interest rate is high for small Business



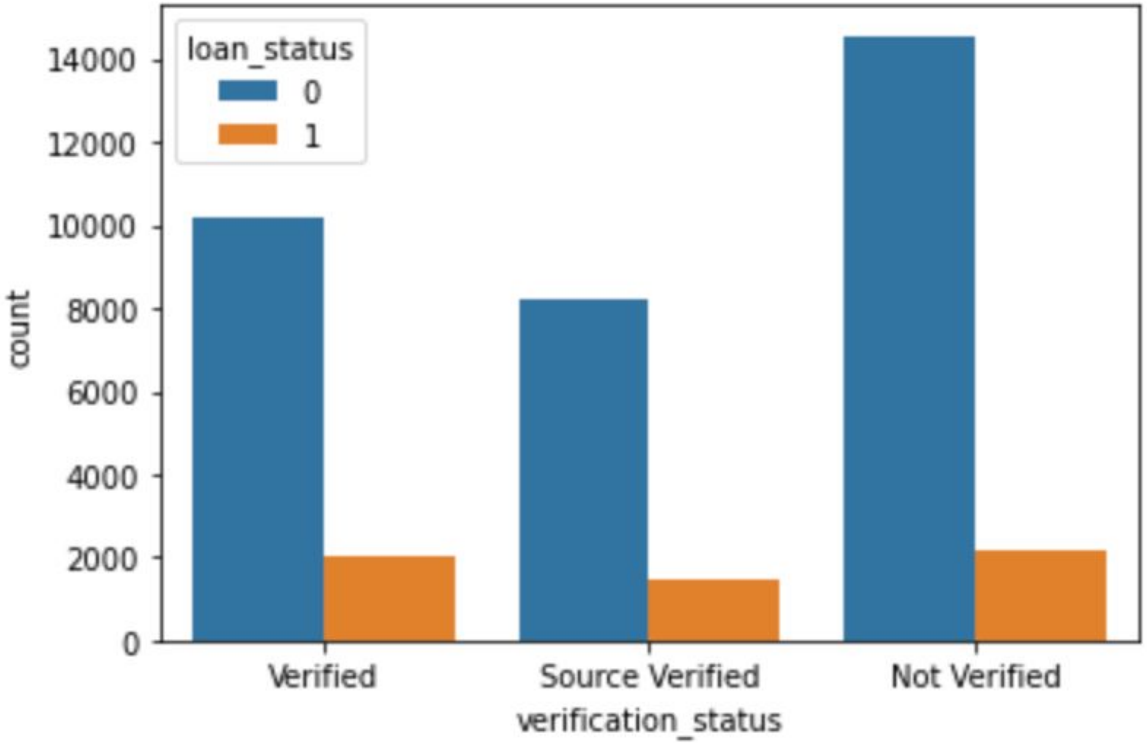
Analysis: the grade is high "A" when the interest rate is less





Analysis: the defaulting is high when the installment is greater than 325

Recommendation: the installments can be maintained within 325 for 60 months term



analysis: not verified verification status has greater chance of defaulting

Recommendation: it is always recommended to do source verification

Key Insights from Analysis

- "annual_inc" has some of outliers
- Most of the borrowers who applied loans have 10+ years of experience
- Most of the borrowers who resides in rented property
- Most of the borrowers applied loan for debt consolidation
- Loan_amnt,funded_amnt,funded_amnt_inv follows same kind of distribution
- the defaulting is high when the installment is greater than 325
- the grade is high "A" for the loan when the interest rate is less
- The interest rate is high for small Business

Key Recommendations

- The borrowers with own house can be granted loan
- it is always recommended to consider multiple parameters when loan is offered above 11000
- The borrowers with annual income greater than 90000 income can be approved loans as they are not defaulting
- The borrowers with dti greater than 7 can be offered loan as they will not default
- Loans on the purpose "renewable_enery","house" and "moving" are profitable as they are less likely to default
- the installments can be maintained within 325 for 60 months term
- it is always recommended to do source verification