

Presentation on Lending Club Case Study

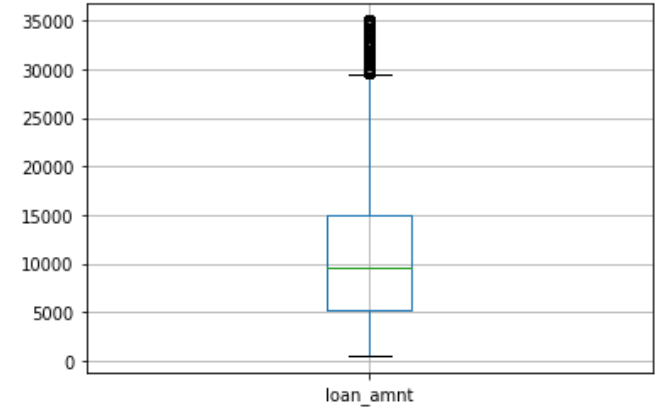
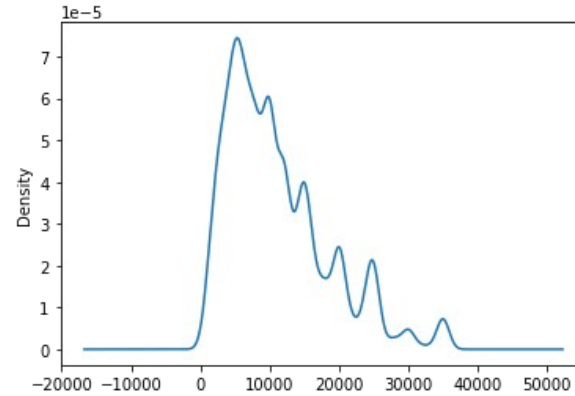
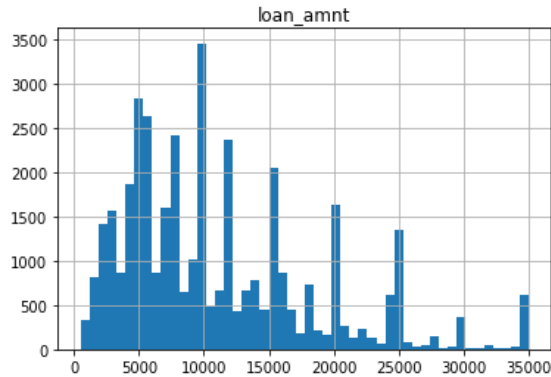
Team Members: Sridhatta Aithal

Analysis on Lending Club Case study Data

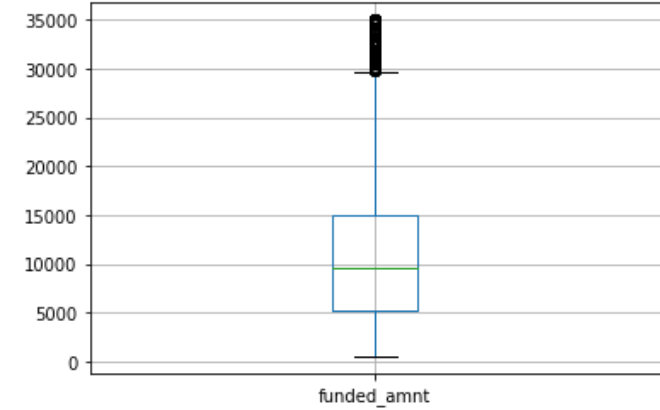
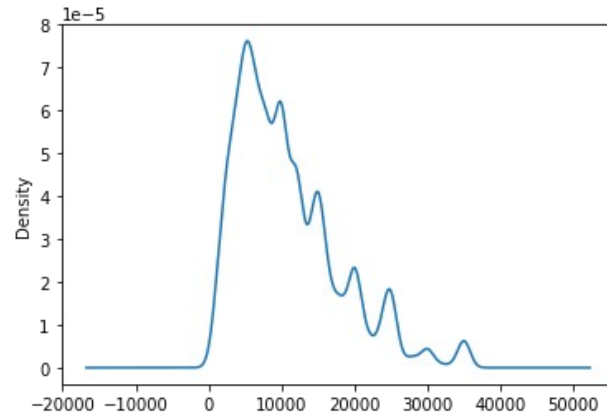
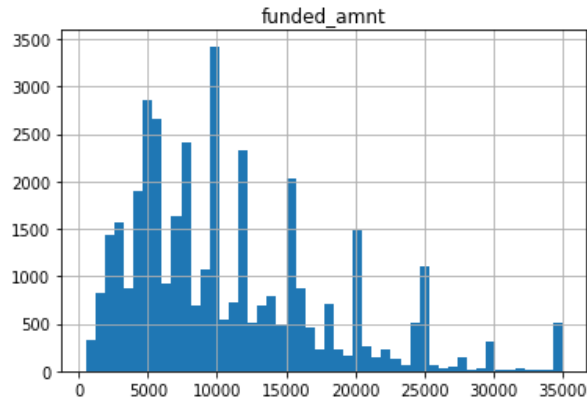
- The dataset provided is to find the objective of rejecting the applicants who can be defaulters in future.
- The pre-process steps of removing the behaviour variables such as delinq_2yrs, earliest_cr_line, inq_last_6mths, open_acc, pub_rec, revol_bal, revol_util, total_acc, out_prncp, out_prncp_inv, total_pymnt, total_pymnt_inv, total_rec_prncp, total_rec_int, total_rec_late_fee, recoveries, collection_recovery_fee, last_pymnt_d, last_pymnt_amt, last_credit_pull_d, application_type are removed as they are not available during the time of loan application.
- The column with all null values are rejected.
- The applicants with loan status of current are removed as they don't contribute to the analysis of the present objective.

- The column of mths_since_last_record had the 92.85% null values hence was removed.
- The column variable of id,member id,pymnt_plan,url,title,initial_list_status,next_pymnt_d,collections_12_mths_ex_med,policy_code,acc_now_delinq,chargeoff_within_12_mths,delinq_amnt,tax_liens,emp_title,zip_code were removed as they have lesser or redundant information for the analysis.

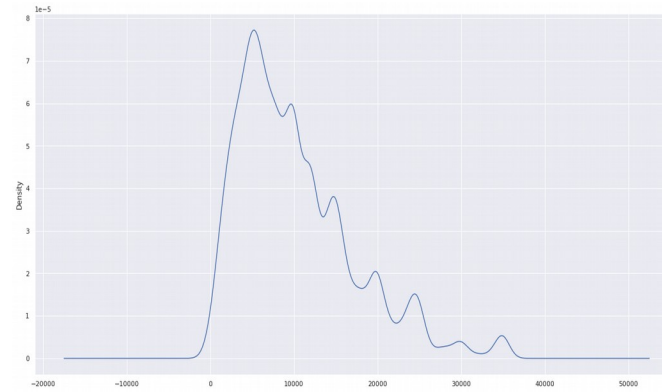
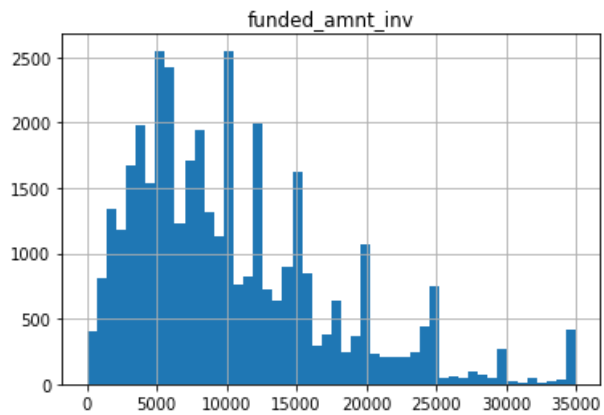
Univariate Analysis



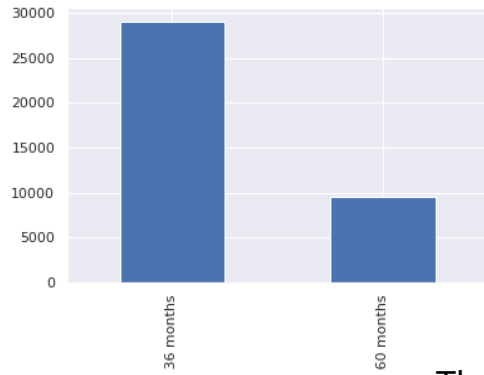
The Maximum amount of loan is drafted within Rs.15000



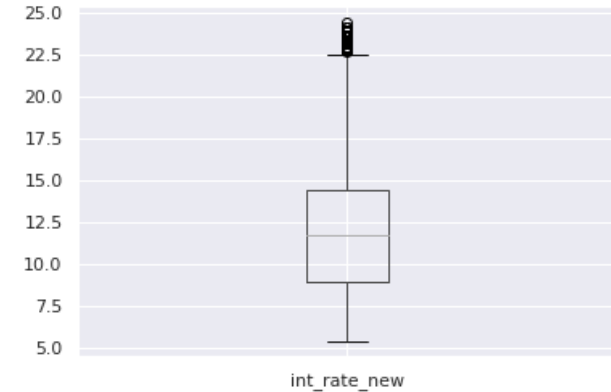
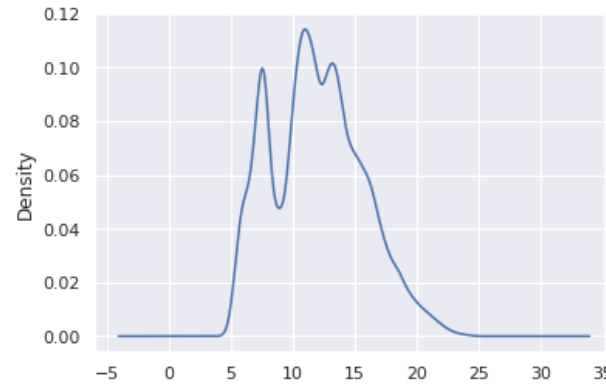
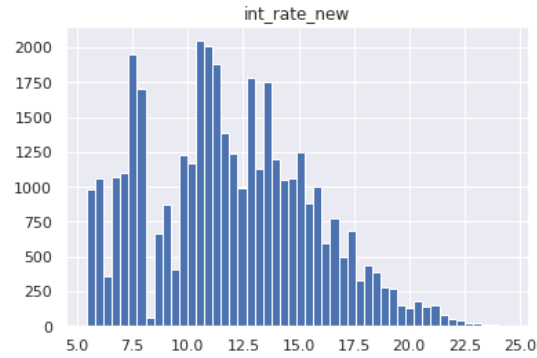
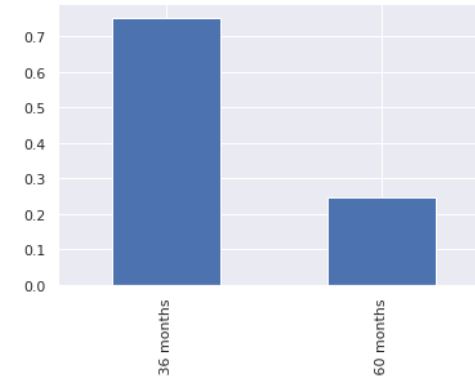
The Maximum amount of loan funded is drafted within Rs.15000



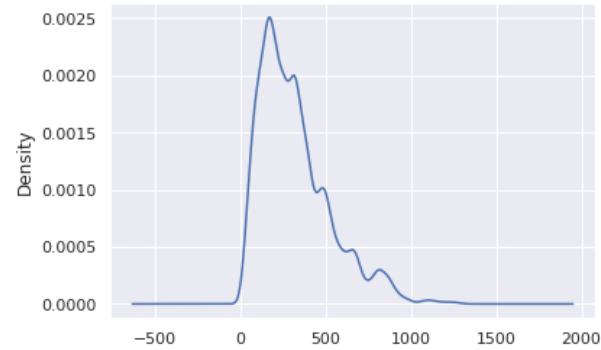
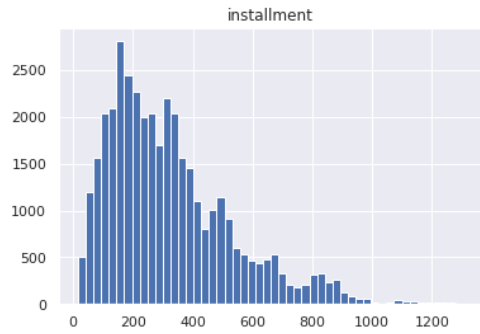
The Maximum loan amount of funded invested is drafted within Rs.15000



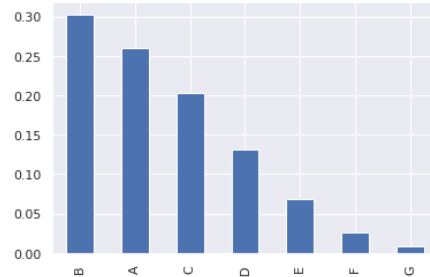
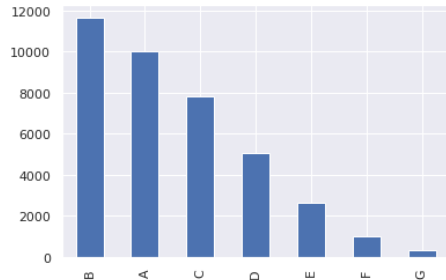
The highest population of term of loan invested is 36 months



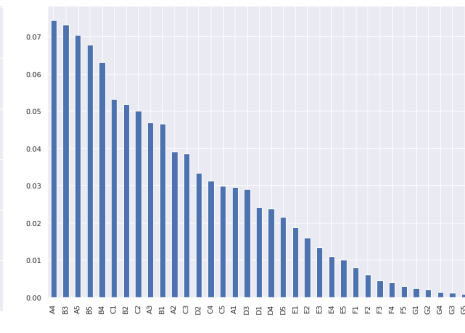
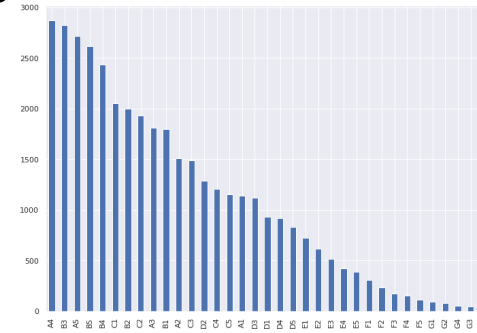
The maximum loan applicant's interest charged is populated within 15%



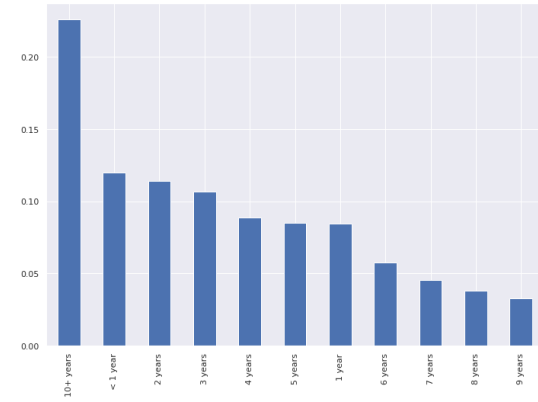
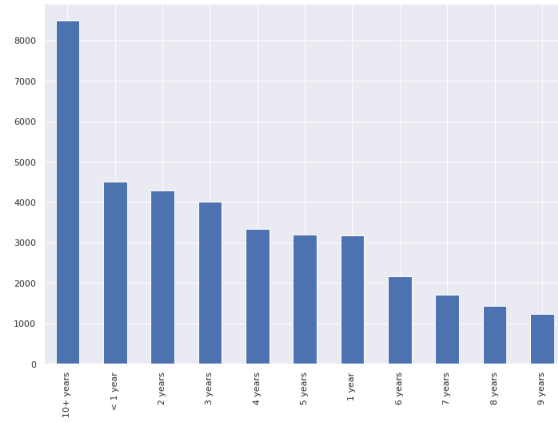
The maximum applicants are paying installments within Rs.500



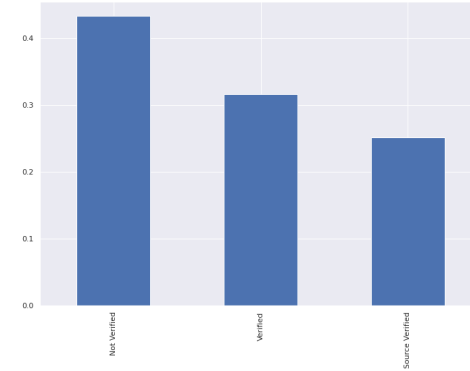
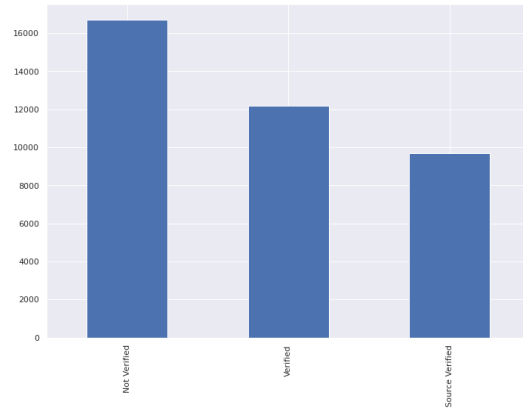
The loan applicants of employment level grade B has the maximum loan followed by A,C



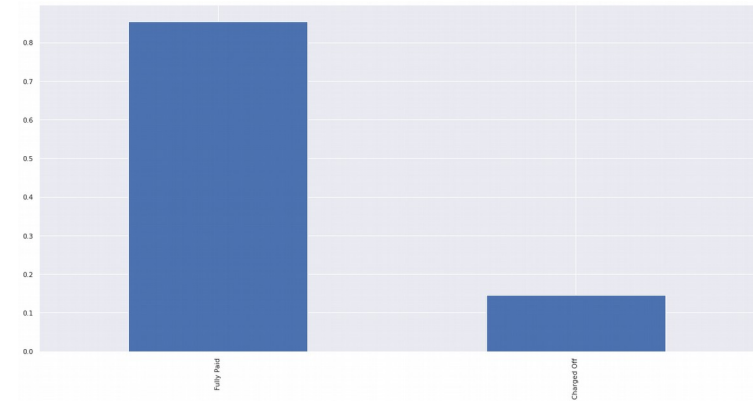
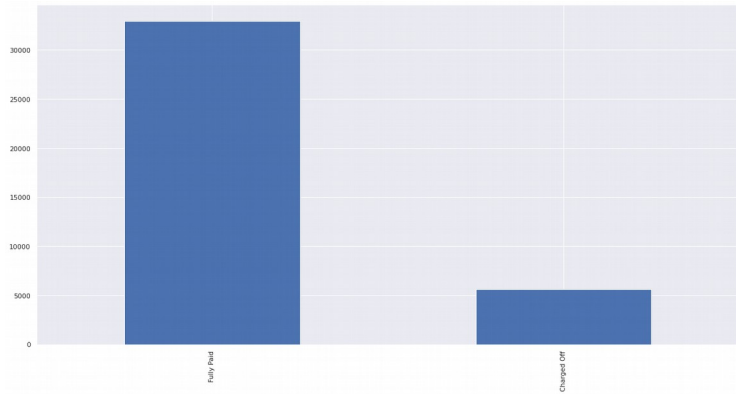
The loan applicants of employment level of sub-grade A4 has the maximum loan followed by B3,A5,...



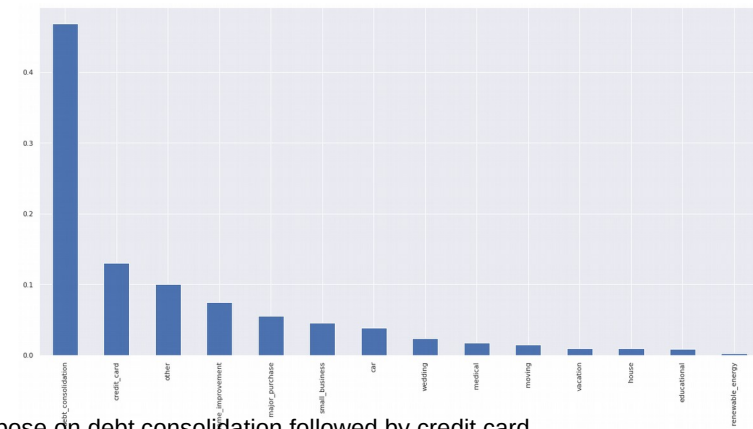
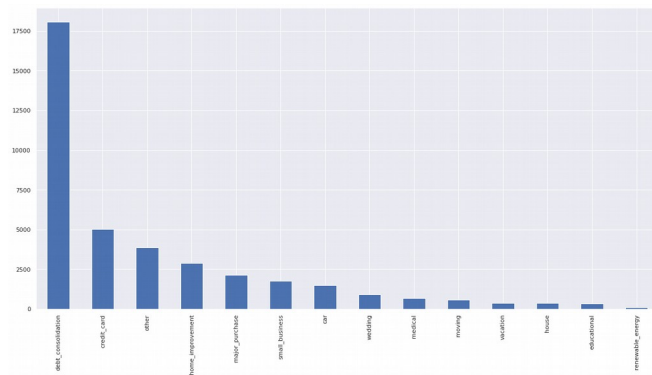
The loan applicants with employee term of 10+ years have the maximum loan followed by <1 year



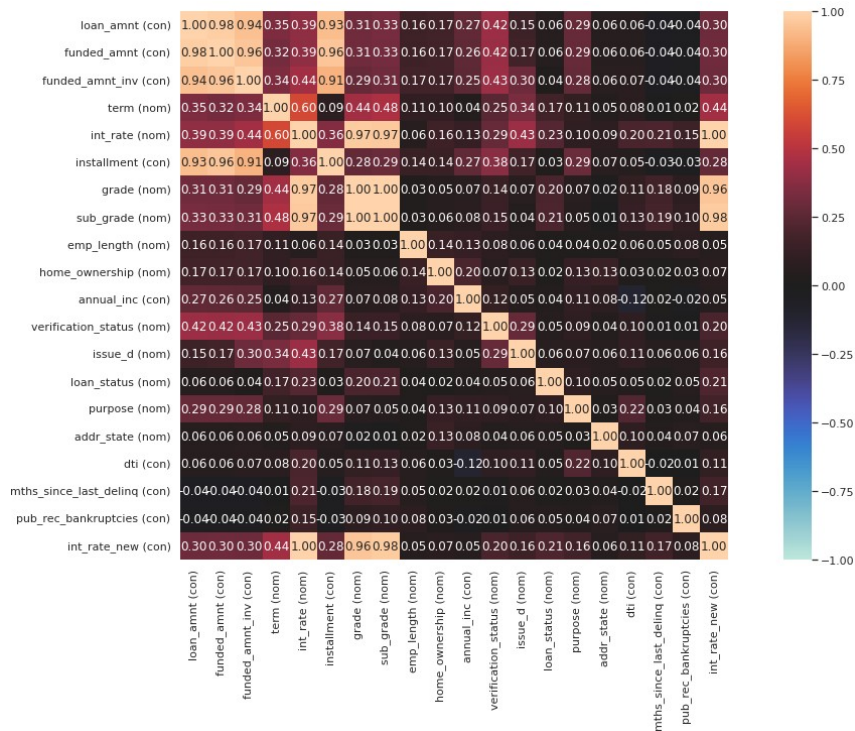
The loan applicants with income verification of Not-verified has the maximum loan applicants



Fully paid loan applicants are in more numbers then charged off(default) category

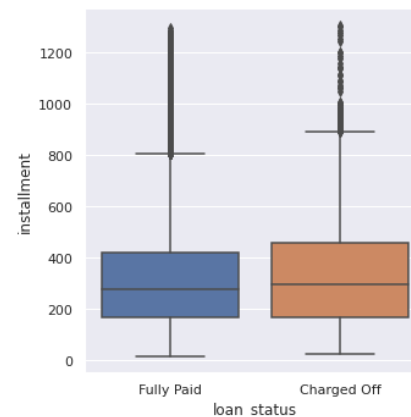
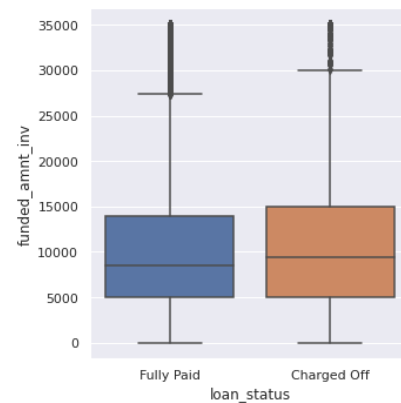
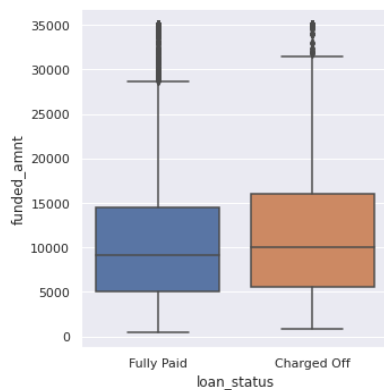
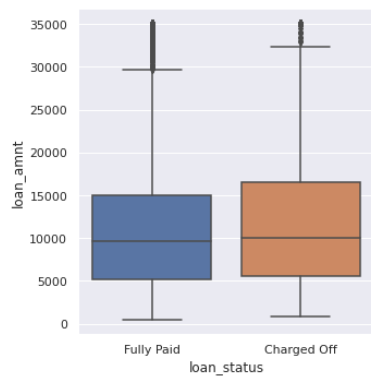


The highest number loan to loan applicants was issued for purpose on debt consolidation followed by credit card

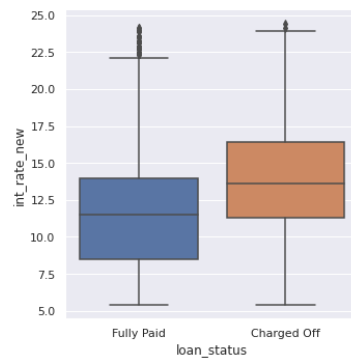


The correlation map tells the overall correlation of variables it is quite high in amount section(loan,funded amount and investment of funded amount) while grade has high correlation with sub_grade and grade and sub-grade has high correlation with interest rate and the loan_status has the highest correlation with interest rate,sub_grade

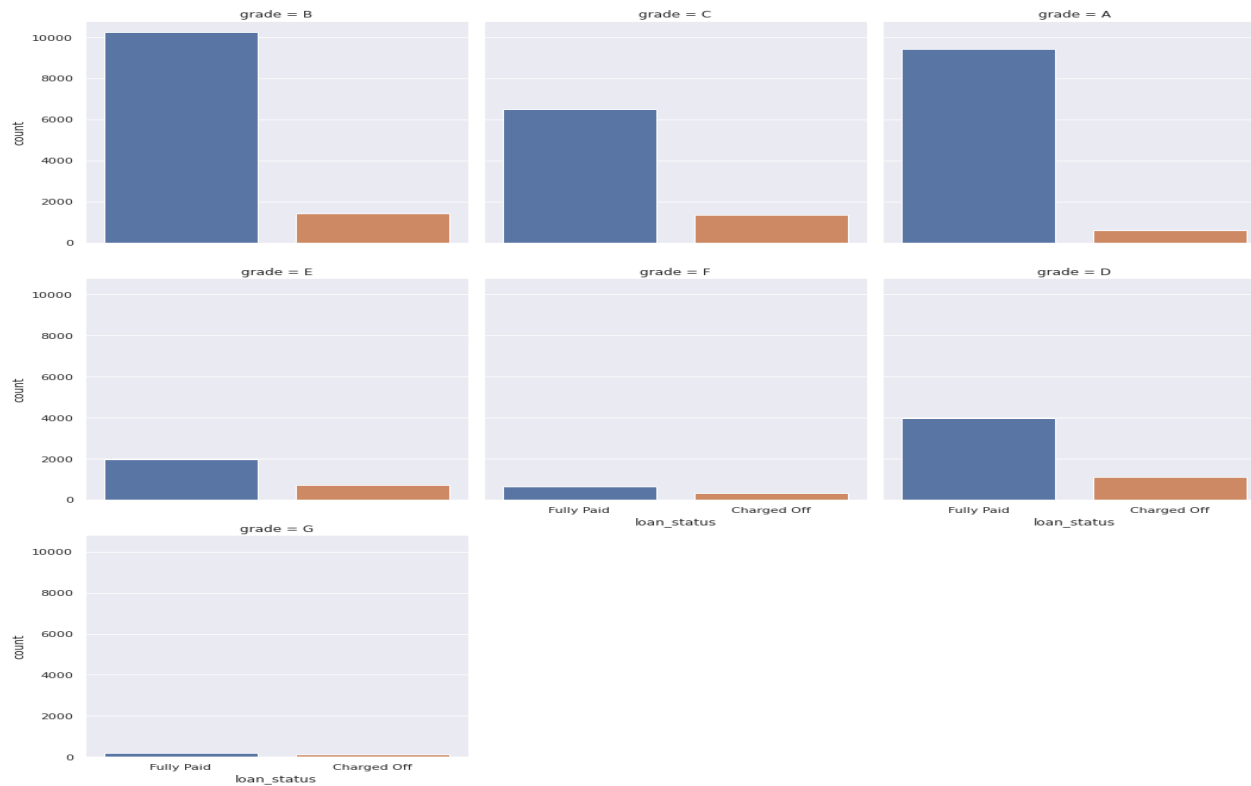
Note: for interest rate variable: int_rate_new is considered



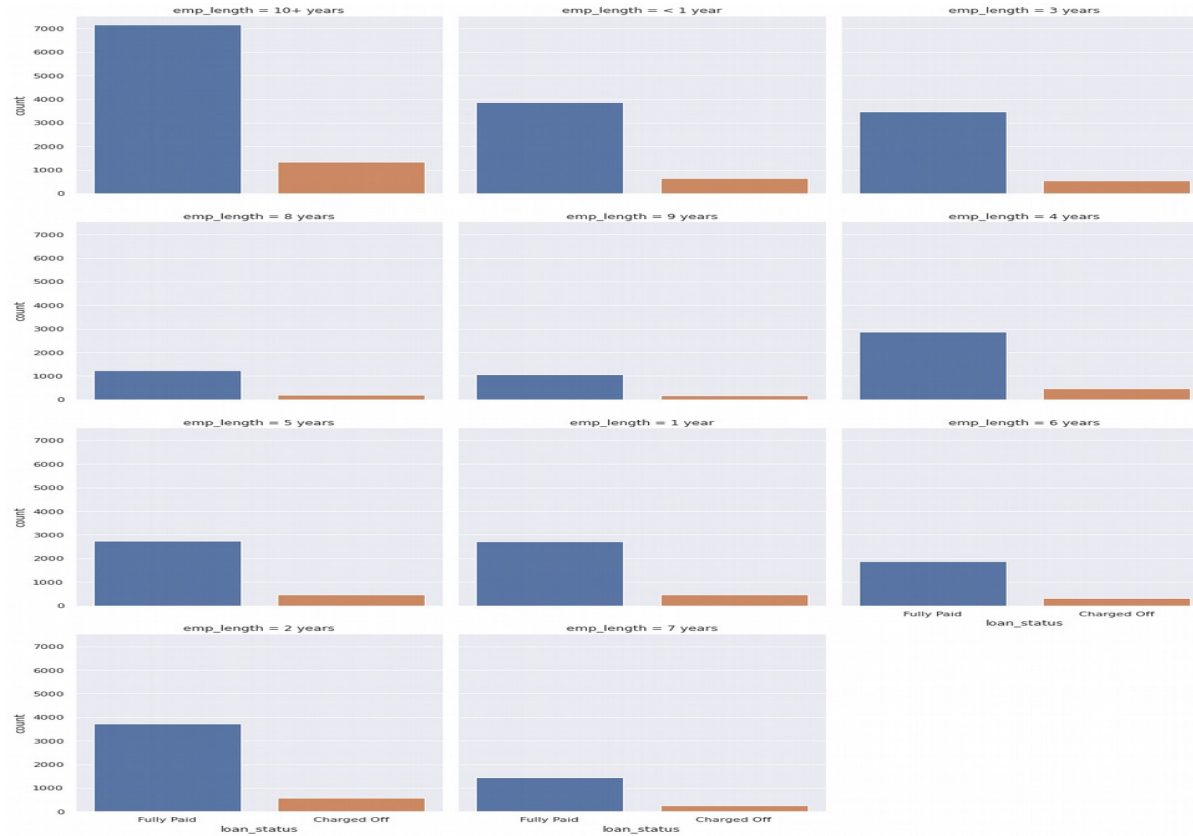
In the above four plots of Box plot the Interquartile range of charged-off i.e. defaulters category is higher then fully paid category



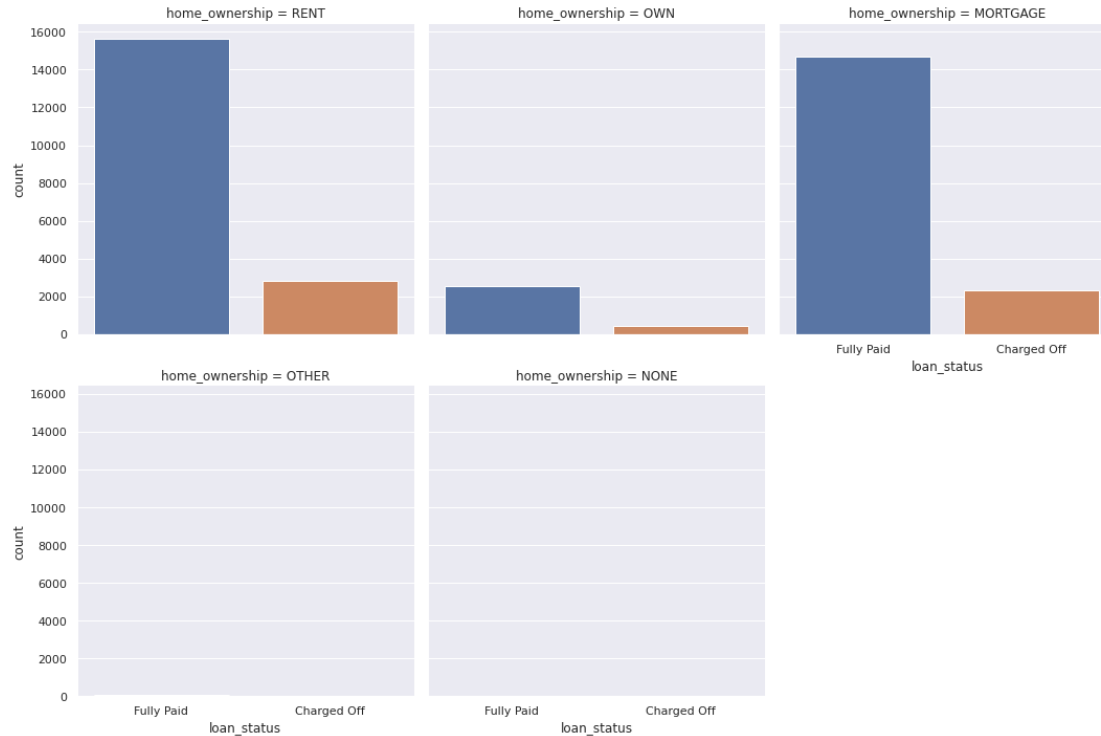
The charged off category(defaulters) has the 25th percentile and 75th percentile greater then fully paid category



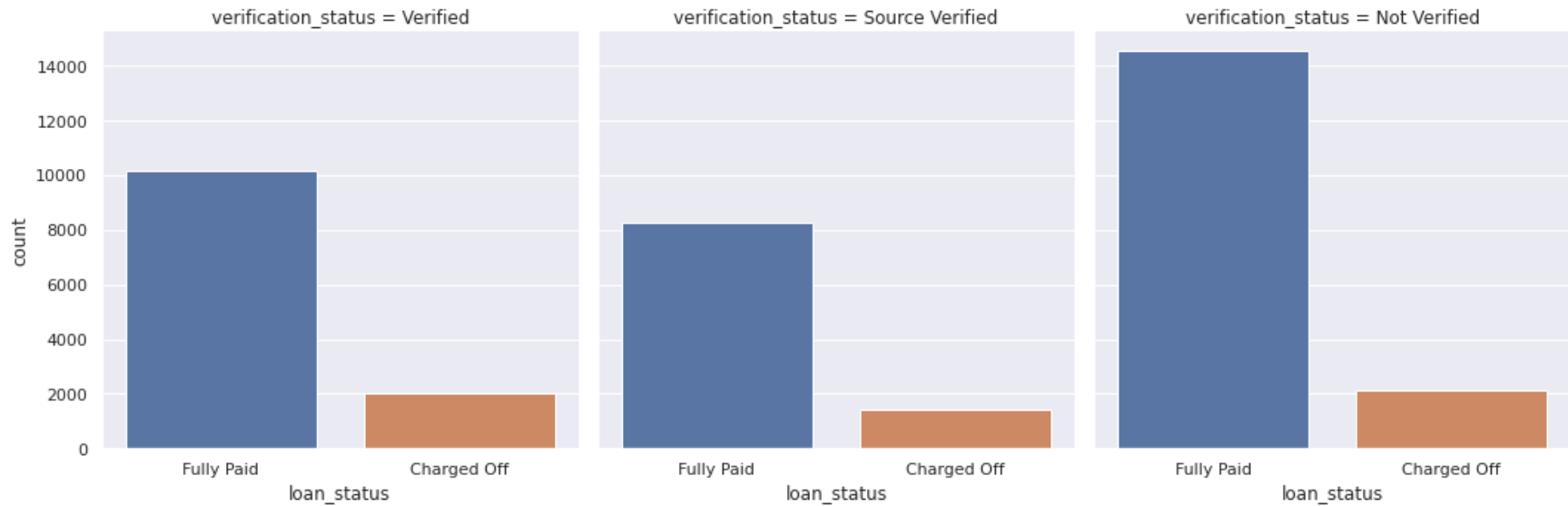
The employee grade B,F of employee has the most of charged-off category(defaulters)



The employee of term 10+years has the most of charged-off category(defaulters)



The most of loan applicants who are of rent category in home ownership has the most of charged-off category(defaulters) while applicants with own house have minimum charged off category.



The loan applicants of source verified has the less charged off applicants.



The loan applicants with the purpose of debt consolidation has the maximum charged off category

Thanks