1

LENDING CLUB CASE STUDY

NIKHIL PATIL
NIMESH JOSHI

BUSINESS OBJECTIVE

The Objective of the case study is to implement EDA techniques for a real world problem and understand the insights to mitigate the credit loss to Lending club

- Identifying the potential candidates who can repay the loan is important, if we reject the person who is an potential borrower and who can repay the loan its an loss to the company .

- If we approve a loan to the person who has very less changes of repaying it, then it's a loss to the leaning club.

With this case study we will be able to pinpoint the applications at risk of defaulting the loans, enabling a reduction in credit loss.

By using the EDA techniques we will be able to understand the driving factor behind the loan defaulters i.e the variables which are strong indicator of default.

CLEANING DATA

Removing columns with NA values

- Mths_since_last_delinq, mths_since_last_record, next_pymnt_d, 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

Removing columns with incomplete or meaning less values

- pymnt_plan, application_type, policy_code, delinq_amnt, tax_liens, initial_list_status, collections_12_mths_ex_med, acc_now_delinq, chargeoff_within_12_mths

Removing Additional Columns

- Dropping these cols 'mths_since_last_record', 'next_pymnt_d' since they have >90% NA values
- Based on this categorical data it seems that "collections_12_mths_ex_med","pub_rec_bankruptcies","tax_liens","acc_now_delinq" are of limited use (only 1 entry)
- Dropping these columns, also dropping "id", "member_id", "url", "title", zip code, emp title, initial list status "verification_status","pymnt_plan","url","desc","title","policy_code","application_type""addr_state", as it is of limited value
- Home Ownership needs to be merged (remove none, other)
- Loan Status with Current Loans Need be removed
- Features are for post loan sanction evaluation

Removing Outliers

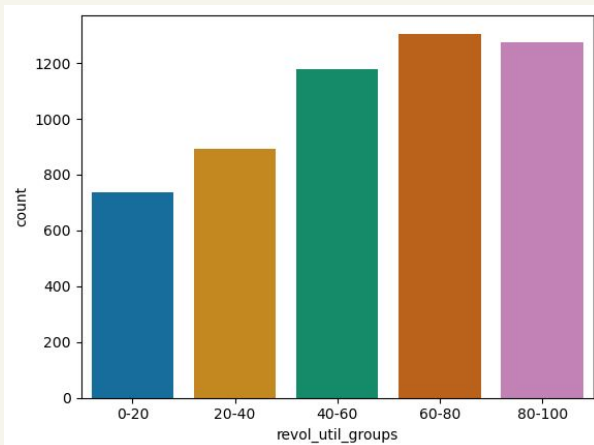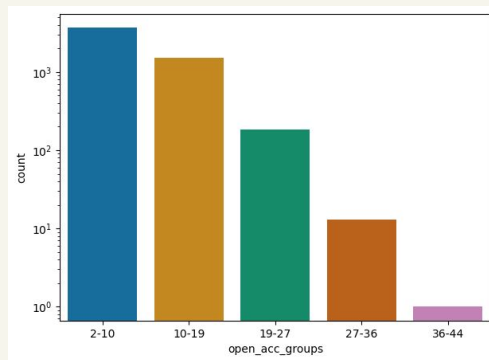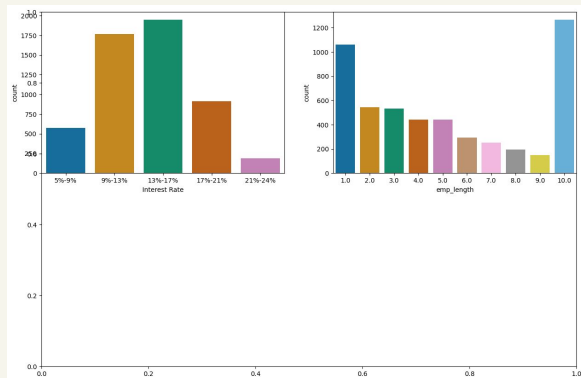- annual_inc, dti, loan_amnt, funded_amnt_inv

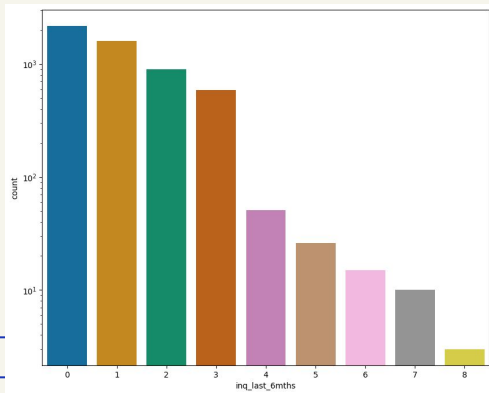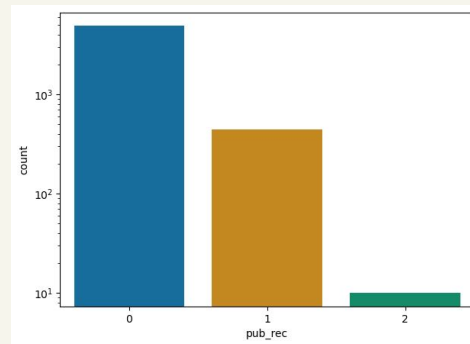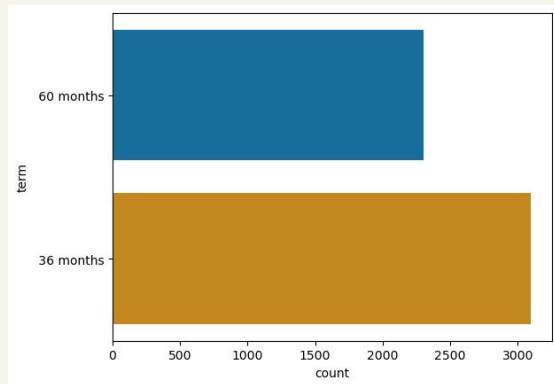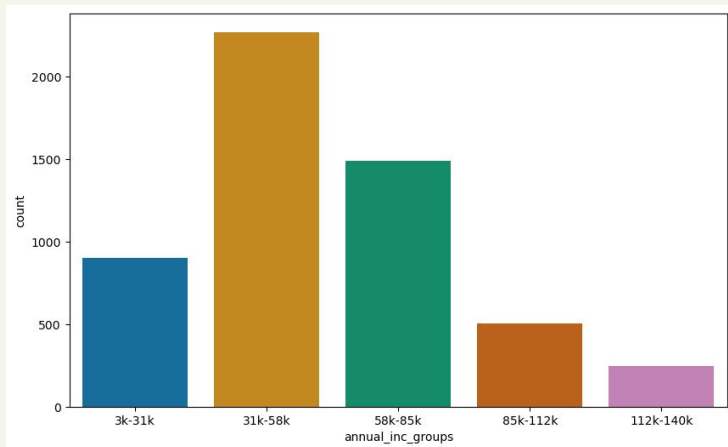# ANALYSIS - 1

Analysis of Defaulter (Charged Off) Data
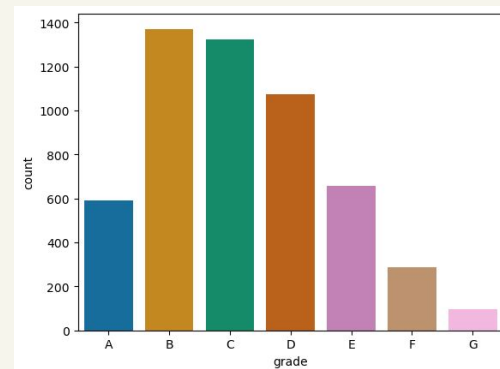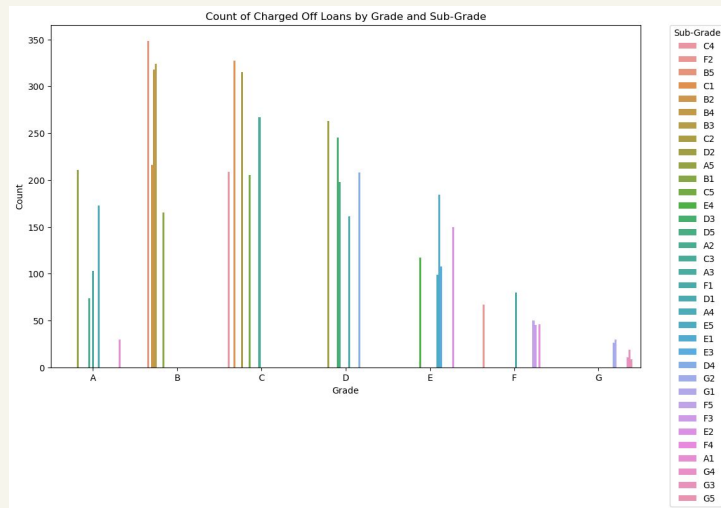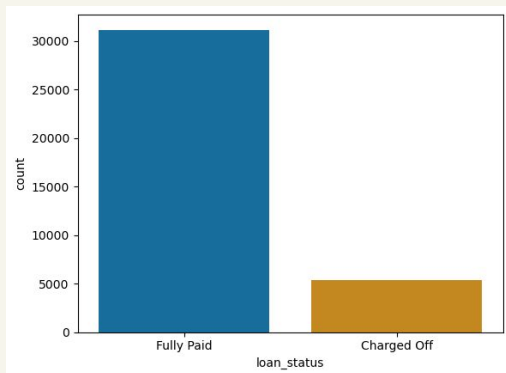
Analysis of Home Ownership Data
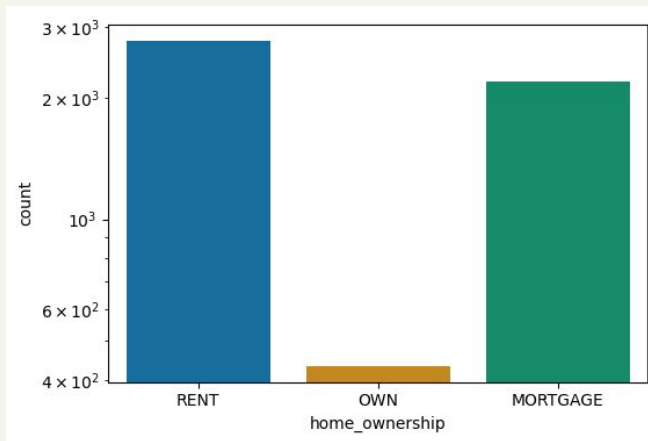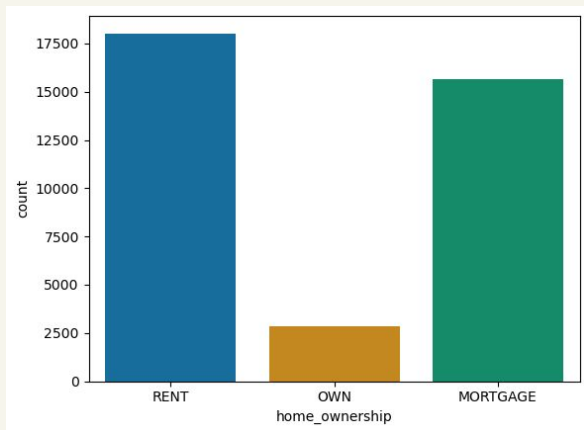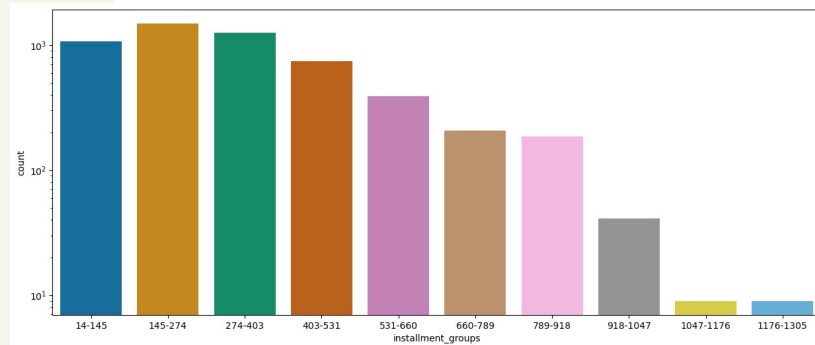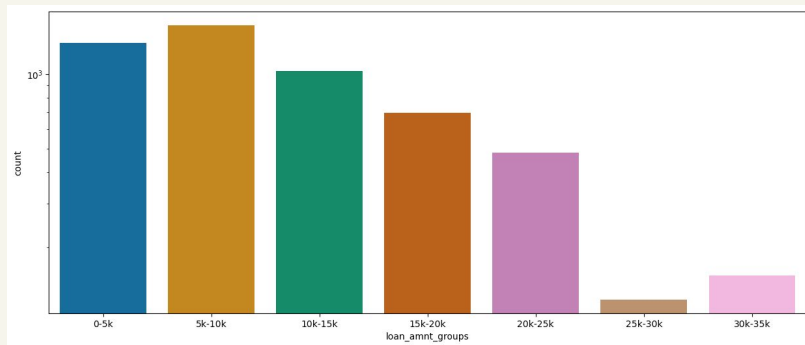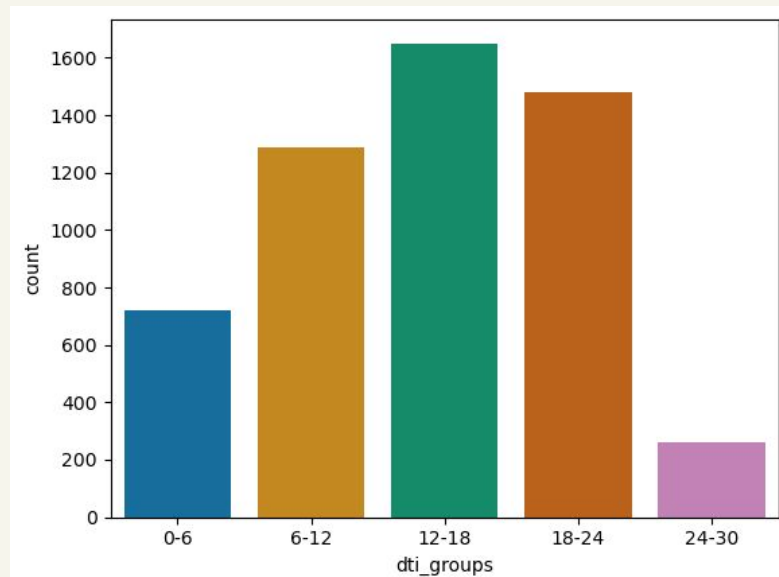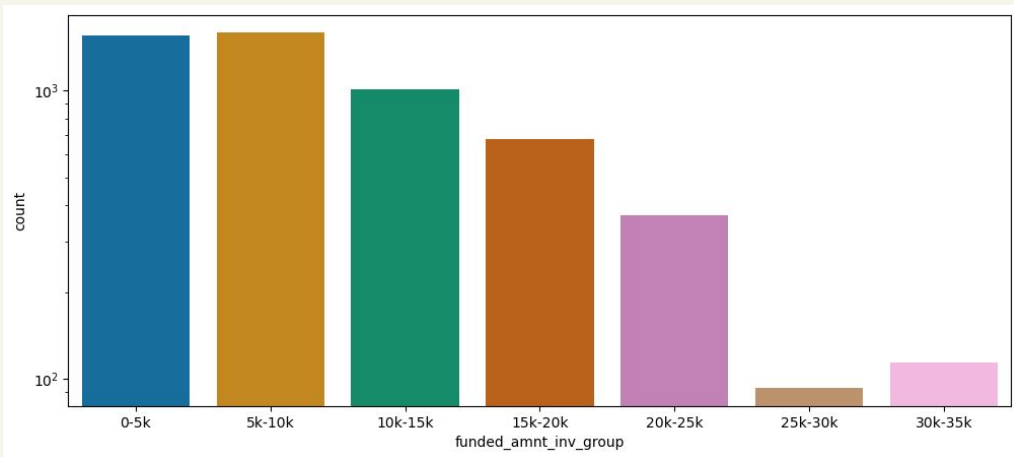
Analysing purpose of loan
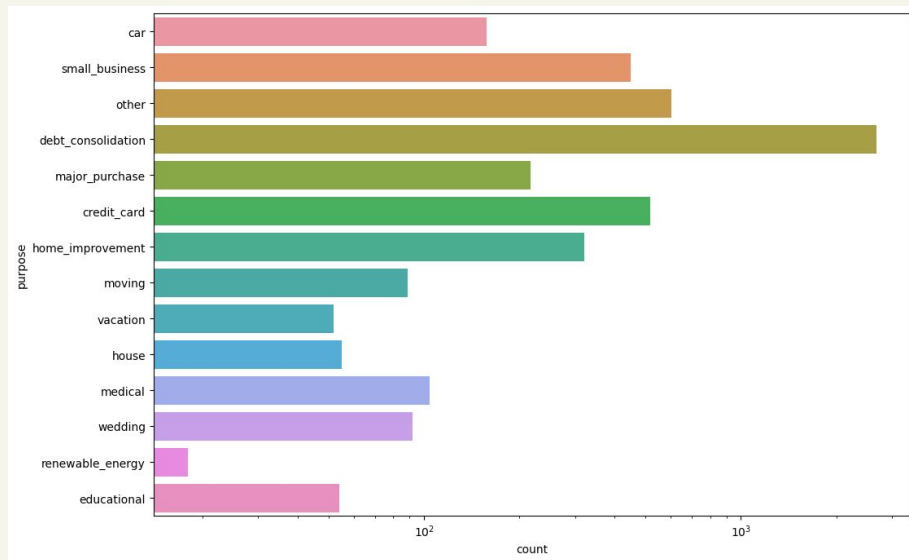
Analysing interest rate
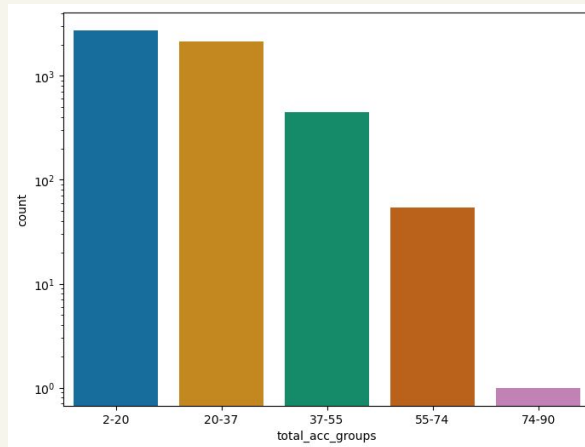
Analysing interest rate
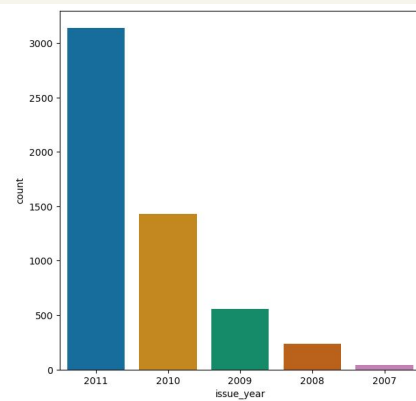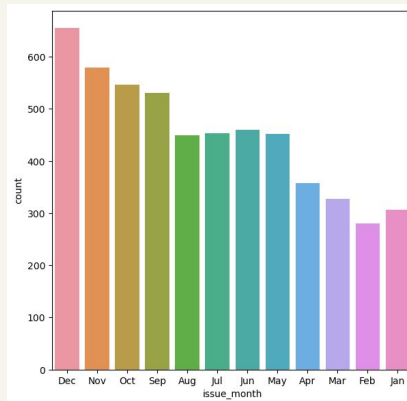
# Analysing installment, dti, loan_amount

OBSERVATIONS - 1

The above analysis concerning the charged off loans for each variable suggests the following. There is a higher probability of defaulting when

- Applicants having house_ownership as 'RENT'
- Applicants who use the loan to clear other debts
- Applicants who receive interest at the rate of 13-17%
- Applicants who have an income of range 31201 - 58402
- Applicants who have 20-37 open_acc
- Applicants with employment length of 10
- When funded amount by investor is between 5000-10000
- Loan amount is between 5429 - 10357
- Dti is between 12-18
- When monthly installments are between 145-274
- Term of 36 months
- When the loan status is Not verified
- When the no of enquiries in last 6 months is 0
- When the number of derogatory public records is 0
- When the purpose is 'debt_consolidation'
- Grade is 'B'
- And a total grade of 'B5' level.
- The 2011 Peak may reflect the financial crisis that was present in the previous years in the USA (2009)
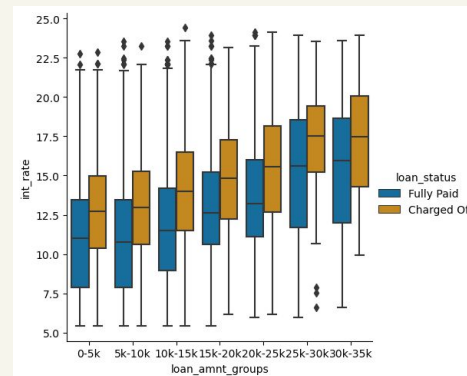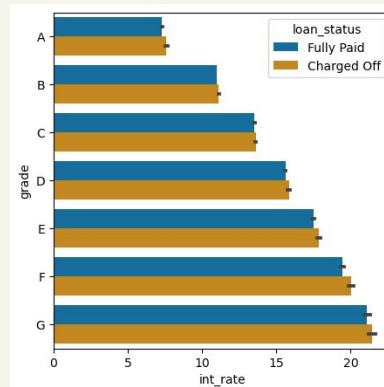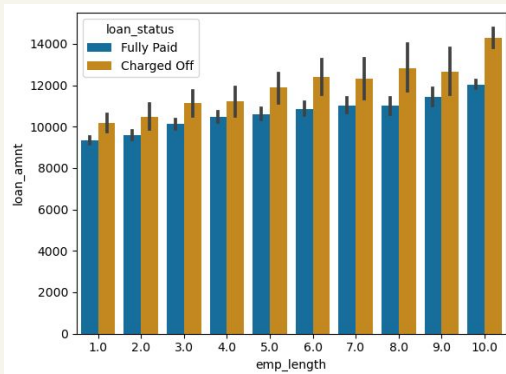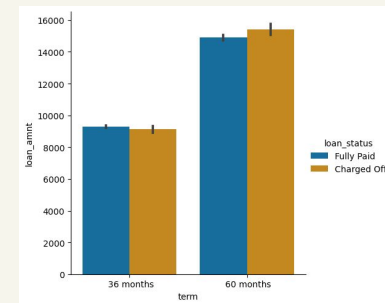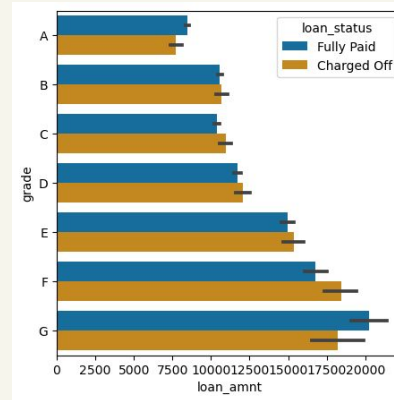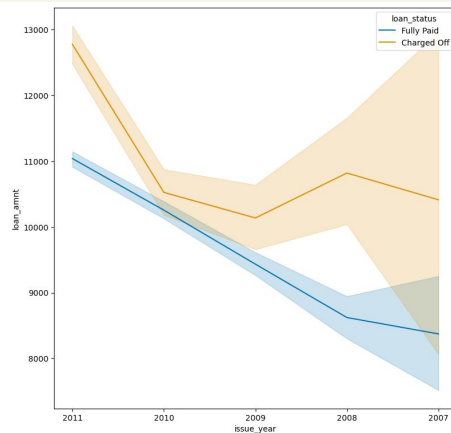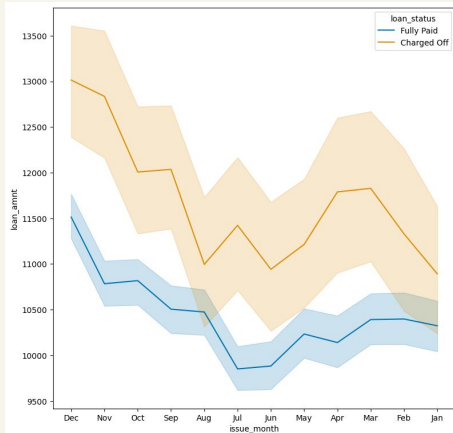
ANALYSIS - 2

Analysing annual income with other columns for more insights
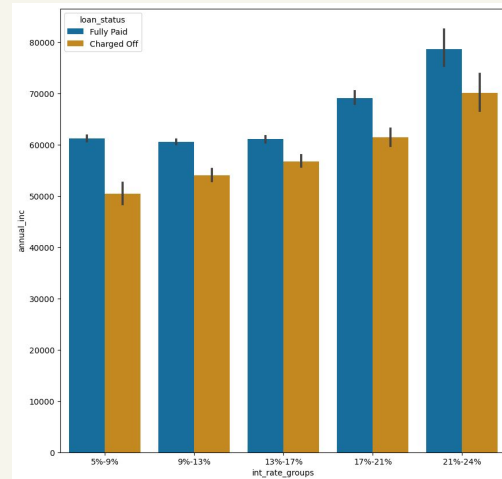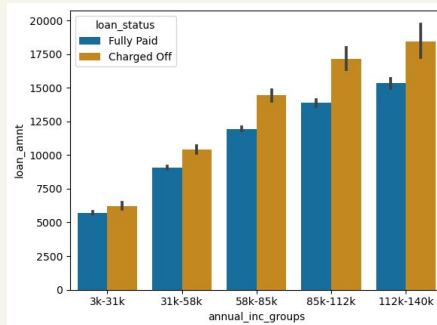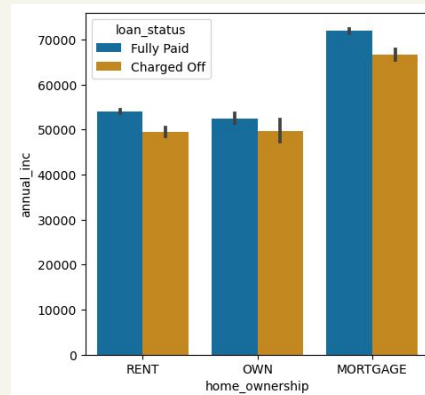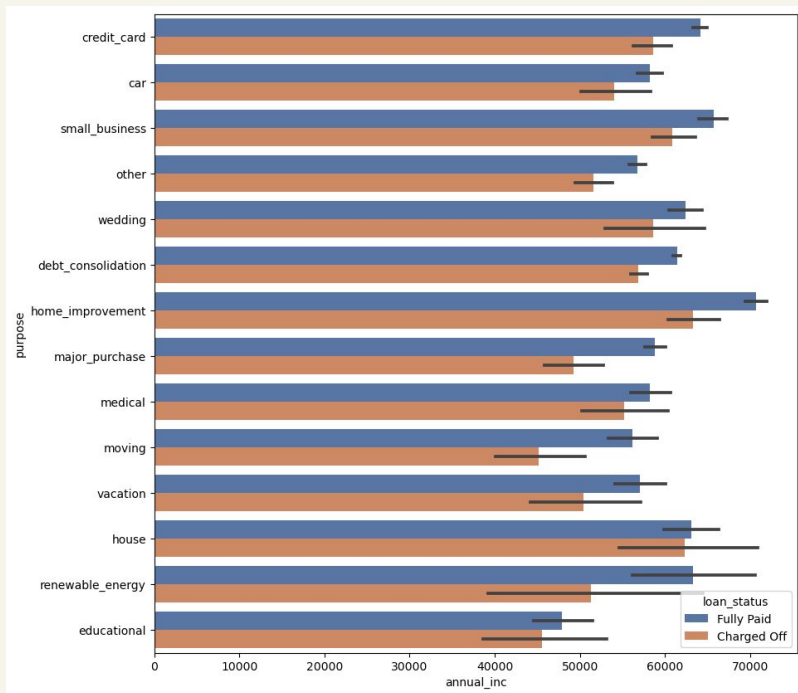
Analysis of loan_amount with other columns for more insights
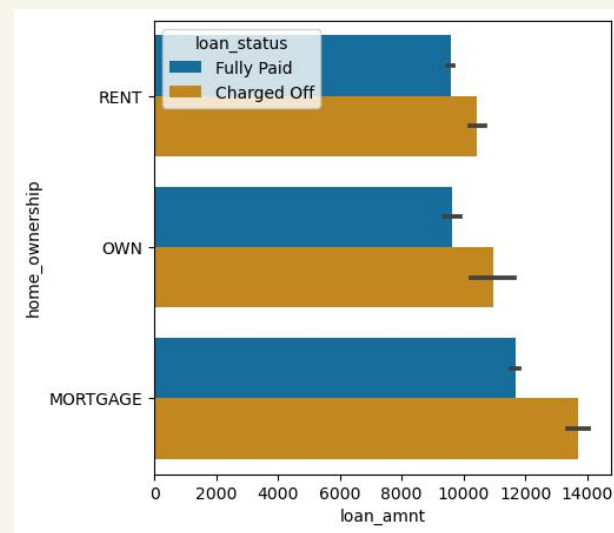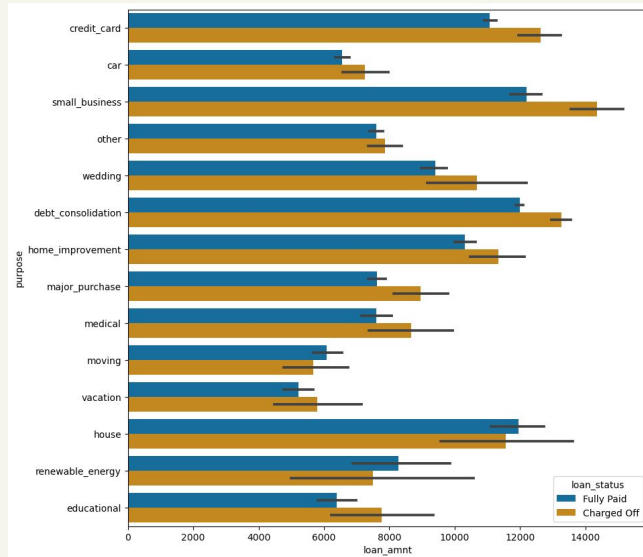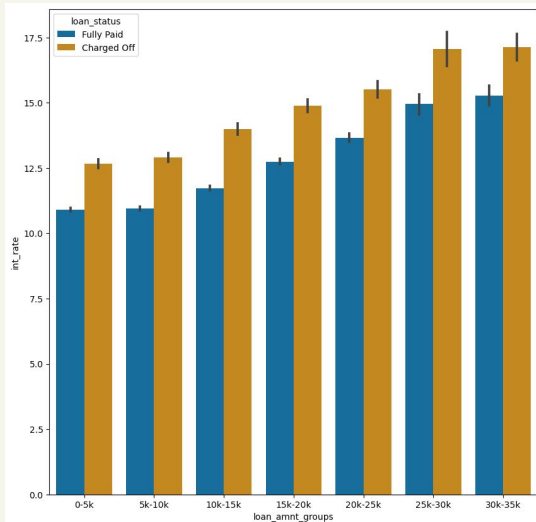
Analysis of loan_amount with other columns for more insights

OBSERVATIONS - 2

The above analysis with respect to the charged off loans. There is a more probability of defaulting when

- Applicants taking loan for 'home improvement' and have income of 60k -70k
- Applicants whose home ownership is 'MORTGAGE and have income of 60-70k
- Applicants who receive interest at the rate of 21-24% and have an income of 70k-80k
- Applicants who have taken a loan in the range 30k - 35k and are charged interest rate of 15-17.5 %
- Applicants who have taken a loan for small business and the loan amount is greater than 14k
- Applicants whose home ownership is 'MORTGAGE and have loan of 14-16k
- When grade is F and loan amount is between 15k-20k
- When employment length is 10yrs and loan amount is 12k-14k
- When the loan is verified and loan amount is above 16k
- For grade G and interest rate above 20%

Thankyou