LENDING CLUB CASE STUDY

DINESH CHAUDHARY
MANASA MARATHE

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.
- The company has to approve loans for the applications received, considering the fact that the invested amount doesn’t go in vain.

Purpose of analysis

- To identify the variables that have more chances of a loan default.
- With this Lending club could identify the potential risks involved when a loan is applied and utilise that to showcase their portfolio.

Data Cleaning

- Before starting the analysis need to drop column with null values and single values columns from the data set has, they have no significance in analysis
- Single Valued columns:
  - 'pymnt_plan', 'initial_list_status', 'collections_12_mths_ex_med', 'policy_code', 'application_type', 'acc_now_delinq', 'chargeoff_within_12_mths', 'delinq_amnt', 'tax_liens

Drop the columns which have no significance to the analysis

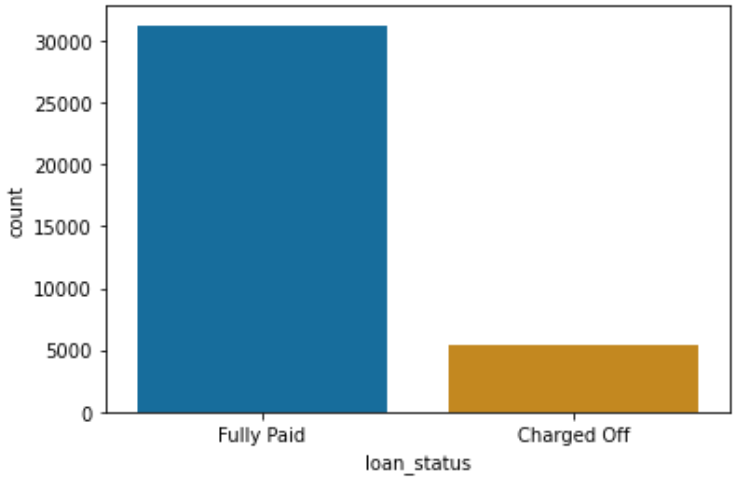
- Some columns are generated post the approval of loan and since we are analyzing the factors contributing to defaulting before approving the loan , we can simply remove these columns.
- Example of post features are delinq_2yrs, revol_bal, out_prncp, total_pymnt, total_rec_prncp, total_rec_int, total_rec_late_fee, recoveries, collection_recovery_fee, last_pymnt_d, last_pymnt_amnt, next_pymnt_d, chargeoff_within_12_mths, mths_since_last_delinq, mths_since_last_record
- Other columns like member_id, id, title , emp_title, desc, zip_code, addr_state, last_credit_pull_d don't contribute to analysis since they are irrelevant.
- funded_amnt is not needed since we only need the actual amount and same is provide by funded_amnt_inv.

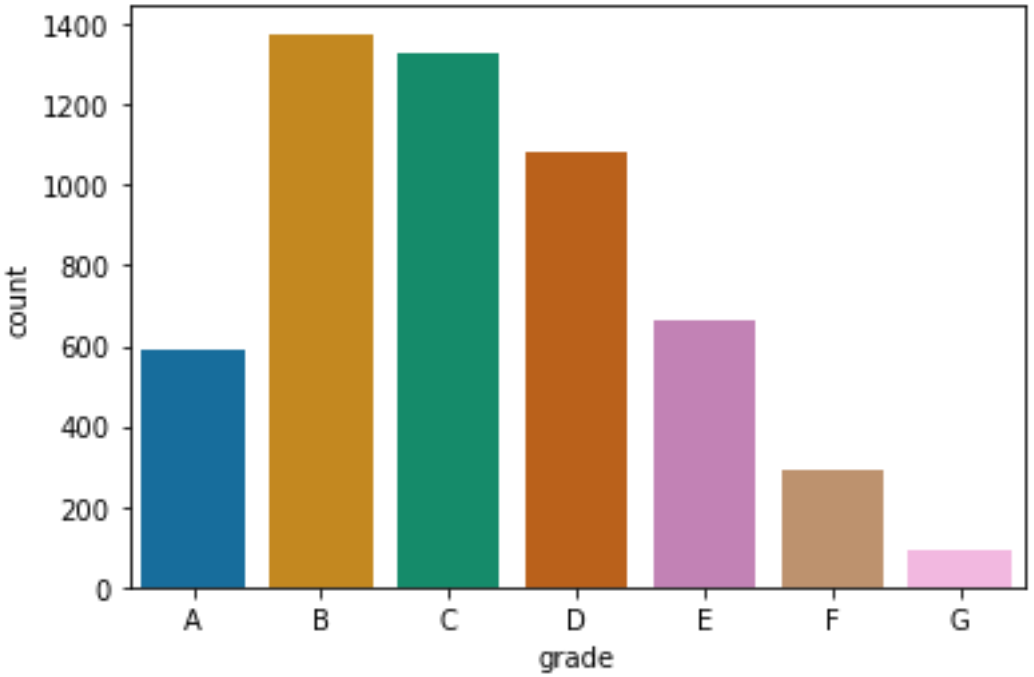
Deriving new columns

- Deriving below columns from existing columns for data analysis
- term_months -- > by removing ‘months’ from column term
- int_rate_percentage -- > by removing ‘%’ from column int_rate
- emp_length_years -- > by removing ‘years’, replacing 10+ with 10 and if <1 replacing with 0.5 from column emp_length
- revol_util_percentage -- > by removing ‘%’ from column revol_util
- issue_d_month, issue_d_year -- > derived from columns issue_d, earliest_cr_line_month, earliest_cr_line_year, earliest_cr_line
- Profil_Loss -- > derived from total_pymnt_inv, funded_amnt_inv

Univariate Analysis / Categorical Analysis of Data

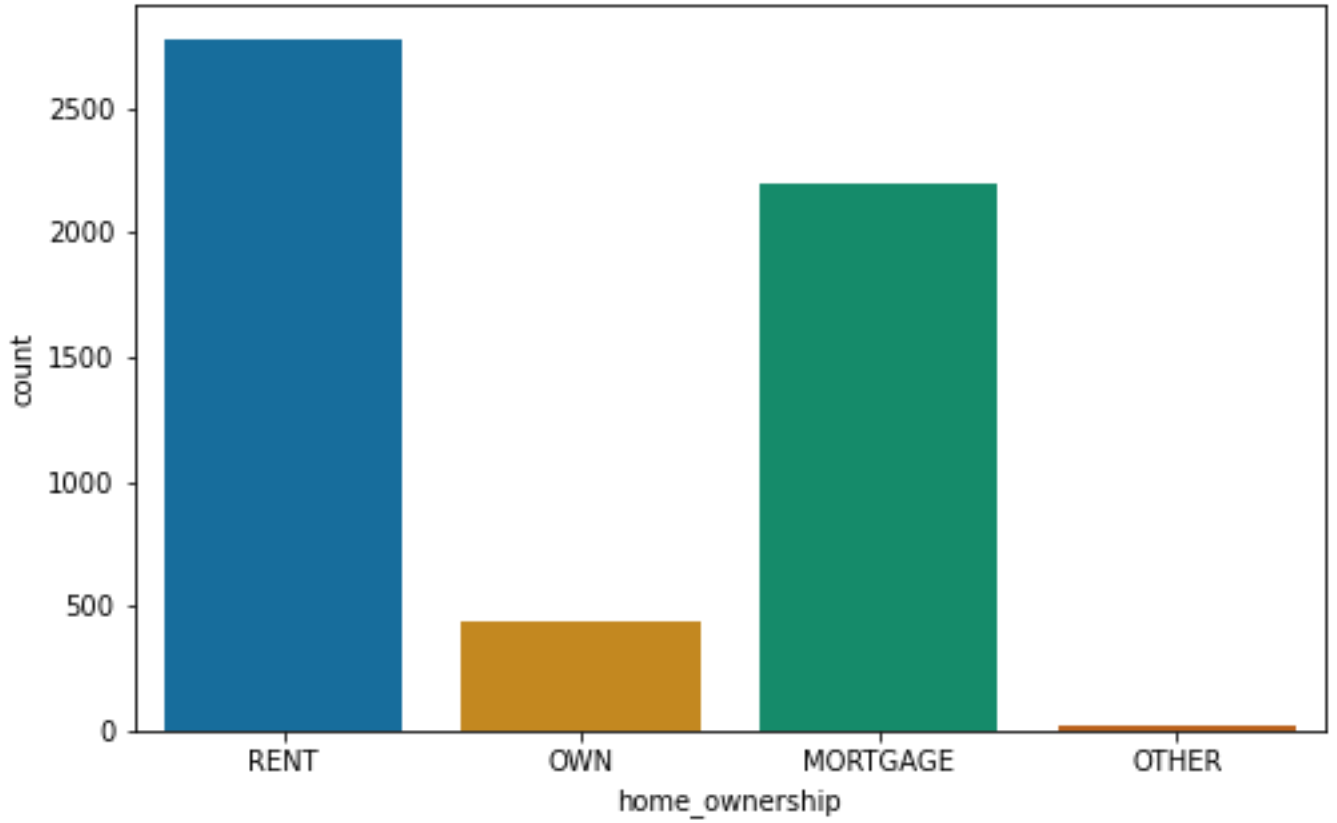
- Since we need to find the parameters that can derive the loan defaulters , let's consider the univariate analysis for loans which have 'Charged Off' loan status.





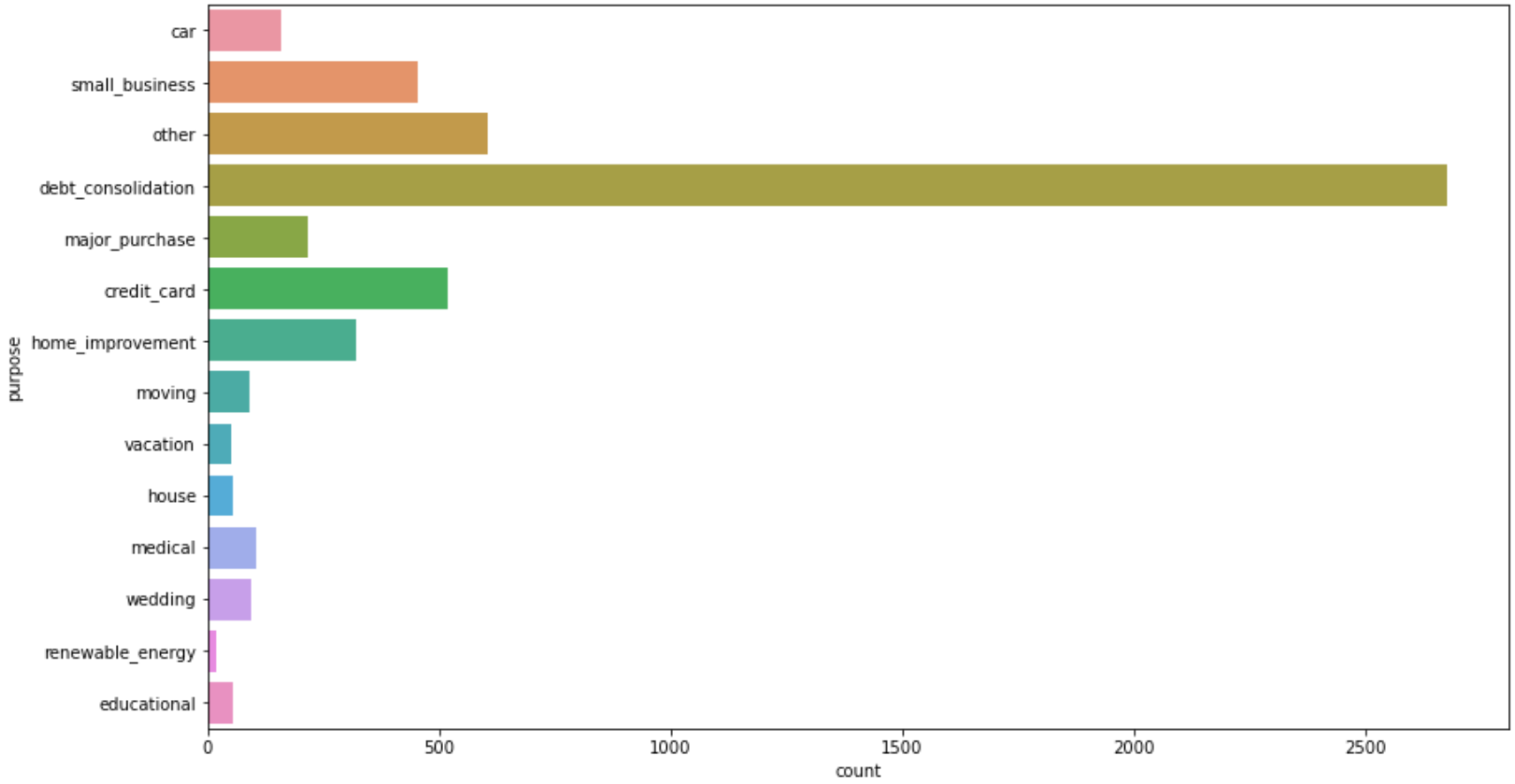
Home ownership analysis

- Inference: There are more chances of defaulting when the applicant is living in rented house



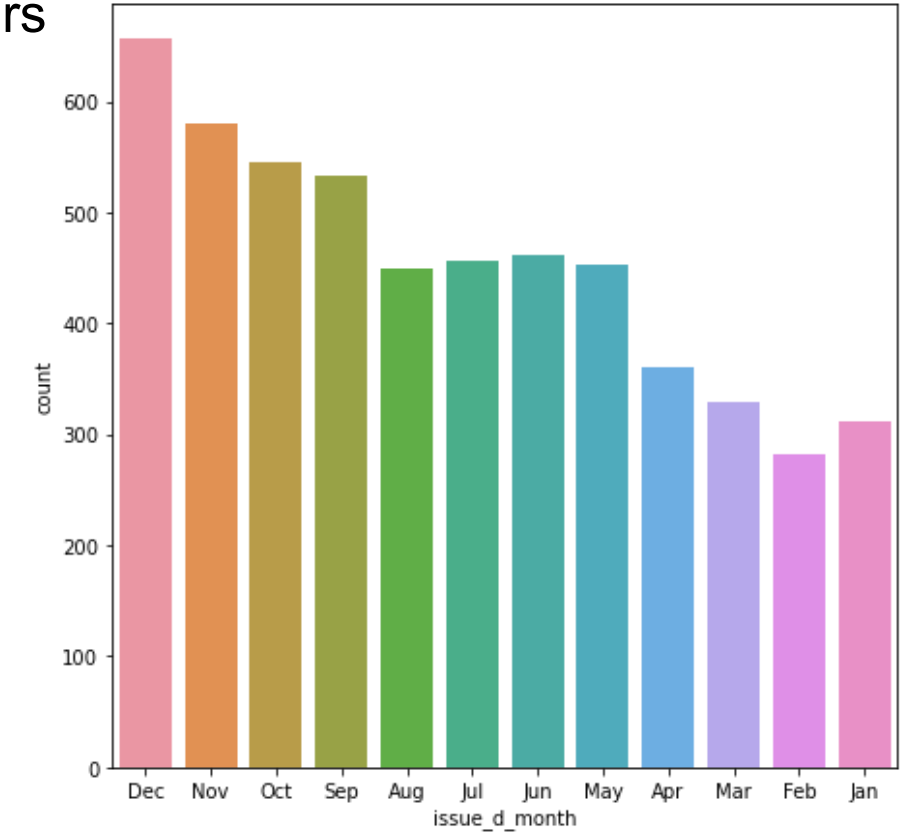
Purpose analysis

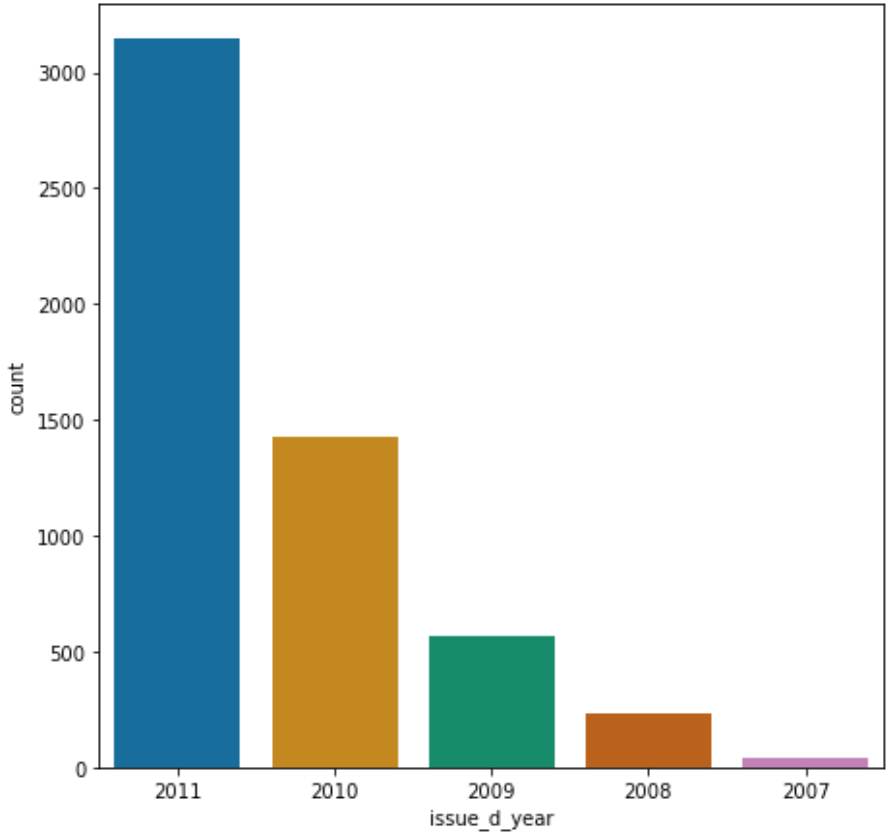
- Inference: Debt consolidation purpose has highest chances of defaults



Analyzing by issued month and year

- **Inference:** Maximum number of defaults have occurred when the loan was sanctioned in Dec. Defaults for loans issued in the year 2011 were also high as compared to other years





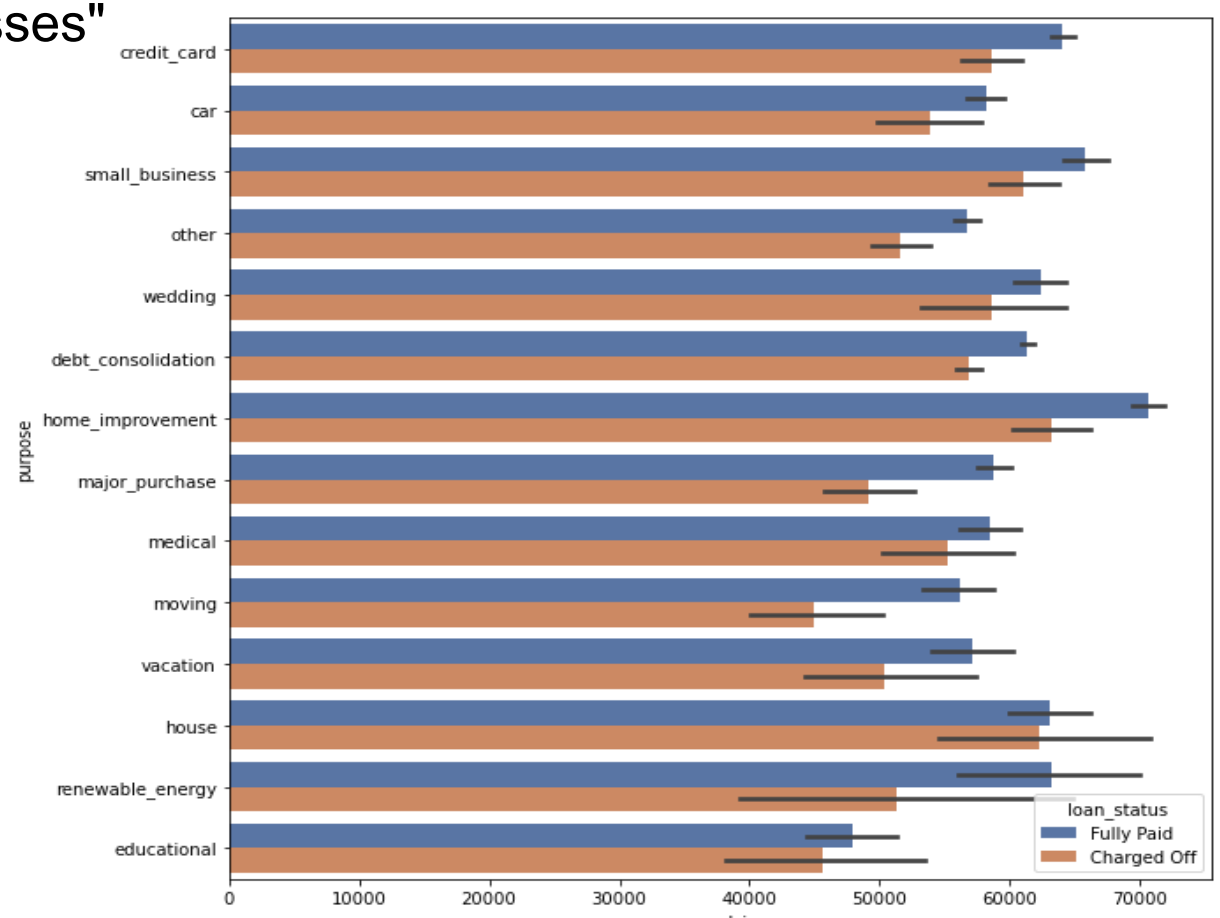
Cumulative Observations

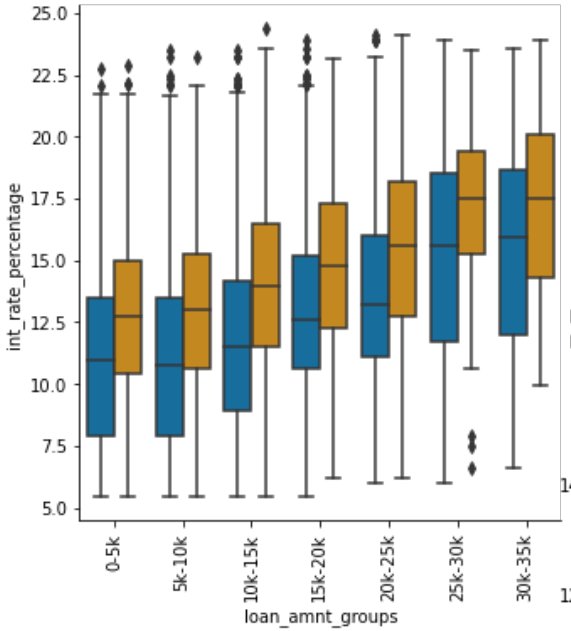
From the plots, we noticed that there are more chances of defaulting when :

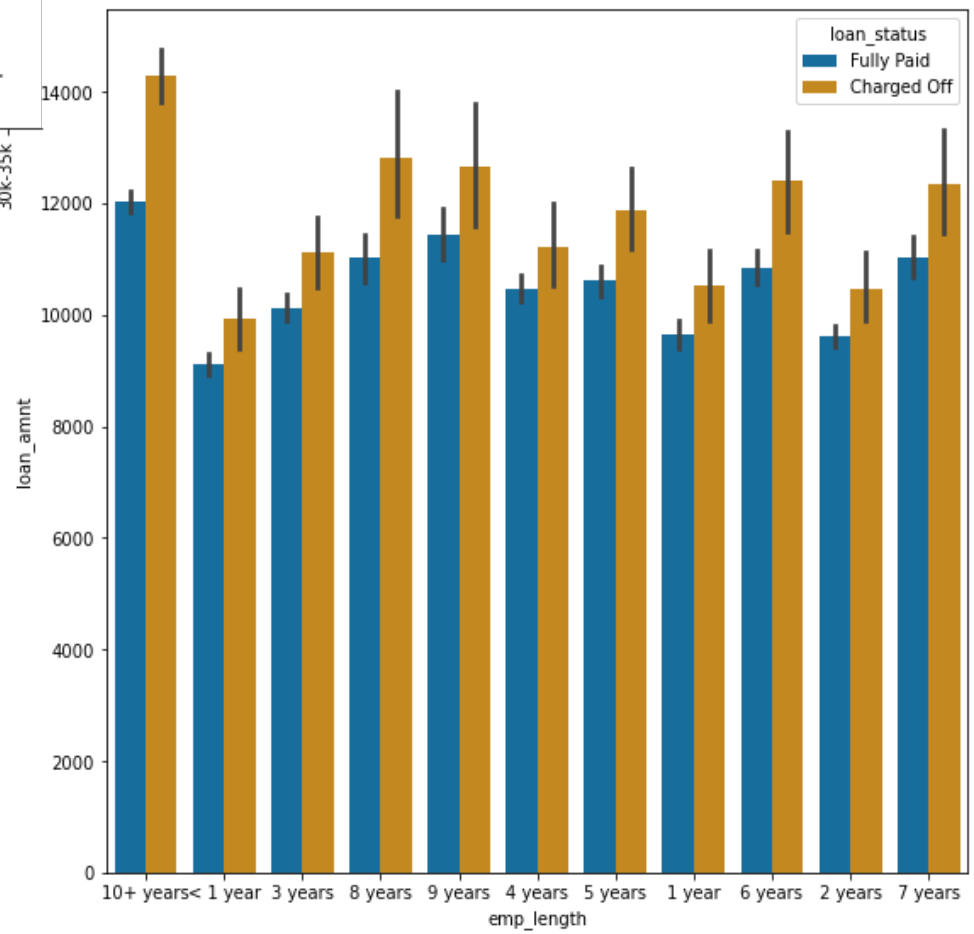
- Applicants having house_ownership as 'RENT'
- Applicants who receive interest at the rate of 13-17%
- Applicants who have an income of range 31201 - 58402
- Applicants with employement length of 10 or more
- When funded amount by investor is between 5000-10000
- Loan amount is between 5429 - 10357
- 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 purpose is 'debt_consolidation'
- Grade is 'B'
- And a total grade of 'B5' level.

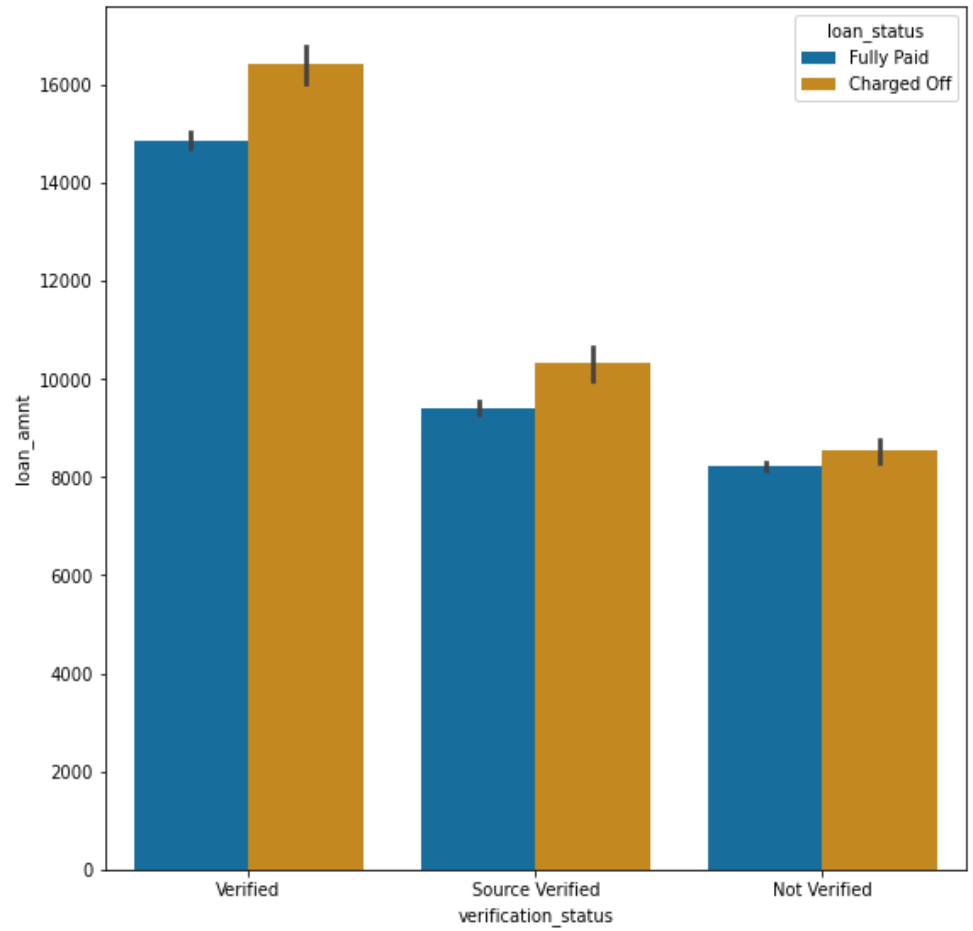
Annual Income vs other columns analysis

- debt_consolidation had highest number of default as seen earlier but the annual income is not highest in that case , instead it is highest in case of home_improvement purpose
- Applicants with higher salary mostly applied loans for "home_improvment", "house", "renewable_energy" and "small_businesses"




int_rate_percentage
loan_amnt_groups
0-5k
5k-10k
10k-15k
15k-20k
20k-25k
25k-30k
30k-35k
loan_status
Fully Paid
Charged Off


loan_amnt
emp_length
10+ years
< 1 year
3 years
8 years
9 years
4 years
5 years
1 year
6 years
2 years
7 years
loan_status
Fully Paid
Charged Off


loan_amnt
verification_status
Verified
Source Verified
Not Verified
loan_status
Fully Paid
Charged Off

Cumulative Observations

With the above plot and analysis of charged off loans. We noticed that the applicant has higher chances 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
- For grade G and interest rate above 20%
- When the loan is verified, and loan amount is above 16k