

Lending Club

Group: 1. Guru Prasad

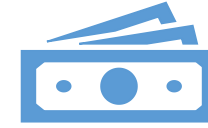
2. Harish Kumar

3. Punitha Girish

4. Sharath Chandra



Lending Club business is frequently exposed to loan defaults risks



Wants to minimize the business risk by minimizing lending to customers with higher default risk profiles



Wants to leverage historic lending data to identify patterns that will aid in decision making

Objective

Resources available

Lending Club Loan
data : loan.csv

Loan data column
dictionary :
Data_Dictionary.xlsx

Approach

1) Data Cleansing

- I. Remove all columns with only one single value and null values; they will not contribute to the analysis
- II. Remove all columns with all null and NaN values

54 columns were removed

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

Approach

2. Data Cleansing

- Formatting
- Identify cells that will need to be imputed
- Identify derived columns

- Format
 - 'int_rate' - remove '%' and covert to float
- Impute Columns
 - 'emp_length' - replace 'n/a' to 'Self -Employed'
- Derived Columns
 - Drive 'issue_month' and 'issue_year' columns from 'issue_d'
 - 'per_loan_amt_to_annual' from 'loan_amnt' and 'annual_inc'

Approach

3. Analysis



Univariate Analysis.

We plot graph for continuous variable and categorical values to find out hidden pattern.



Bivariate Analysis

We took both continuous and categorical values and analysis with target column and figure out relation between them



Correlation Analysis

We designed correlation matrix and figure out co relation among columns

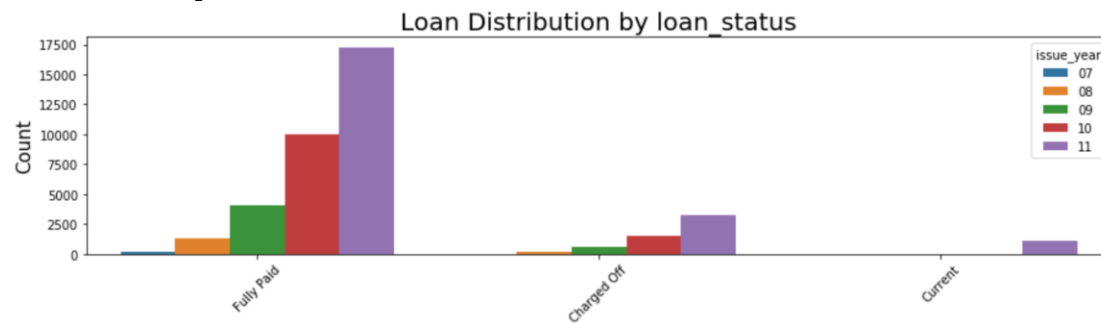


Hypothesis Testing

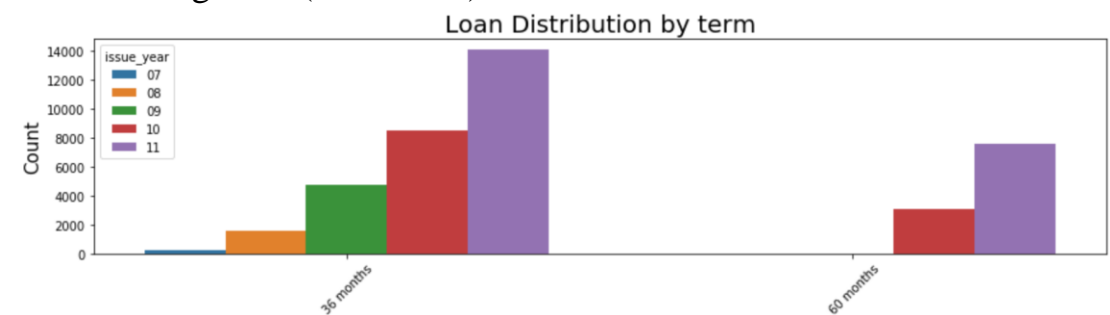
We did EDA so hypothesis testing is not required.

Univariate Analysis

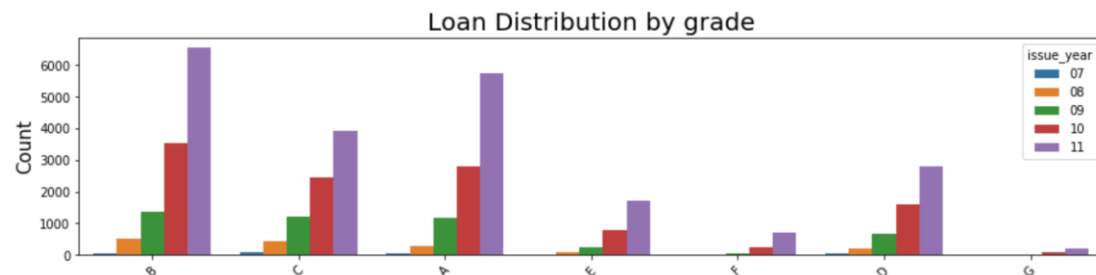
- Loan recovery has been relatively better in 2011 compared to other years



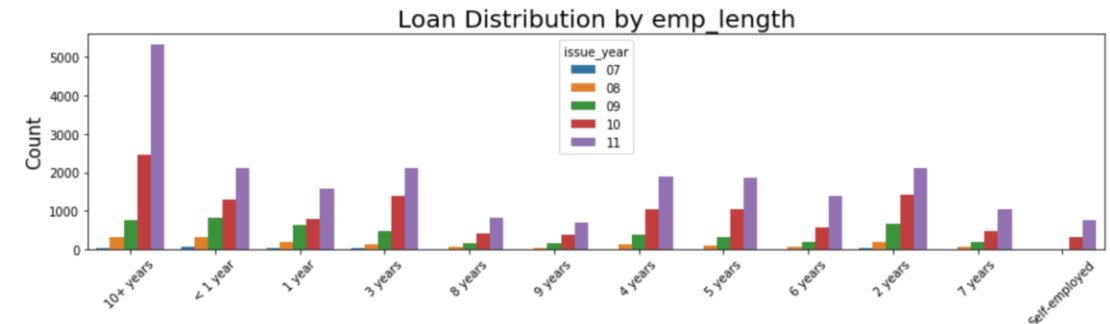
- We are seeing a trend of applicants seeking short term loan, from 2010 long term (60 months) loans were issued



- Clients with higher grade were given the maximum loan

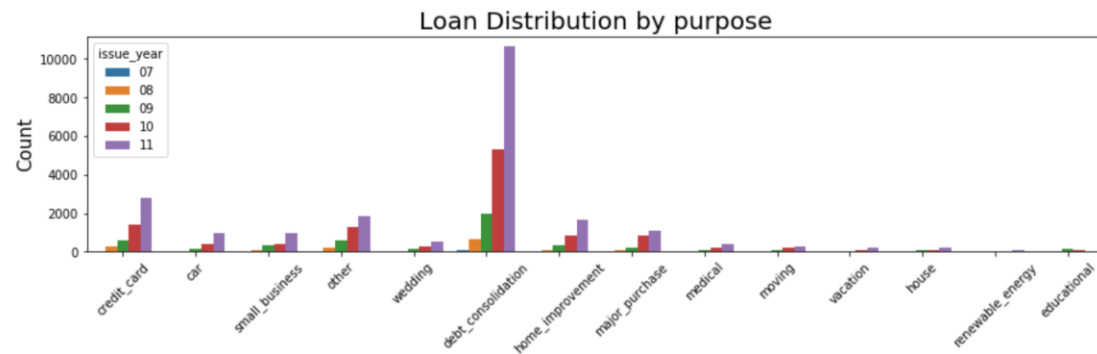


- Applicants employment stability has been a consistent influencer for loan approval. 2011 is seeing a higher approval rate for applicants with higher employment tenor

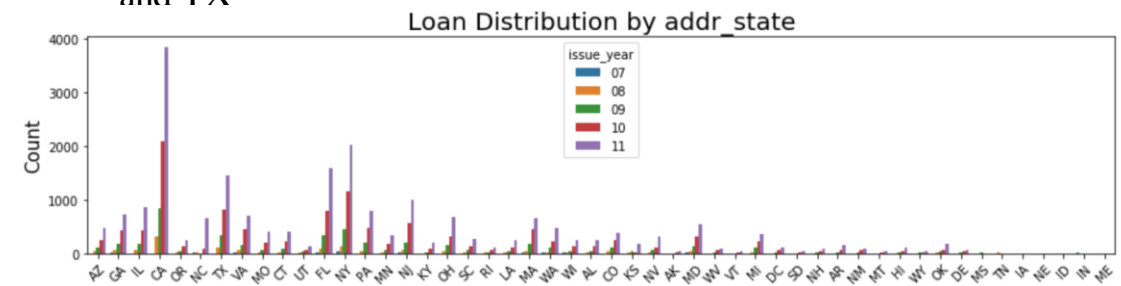


Univariate Analysis

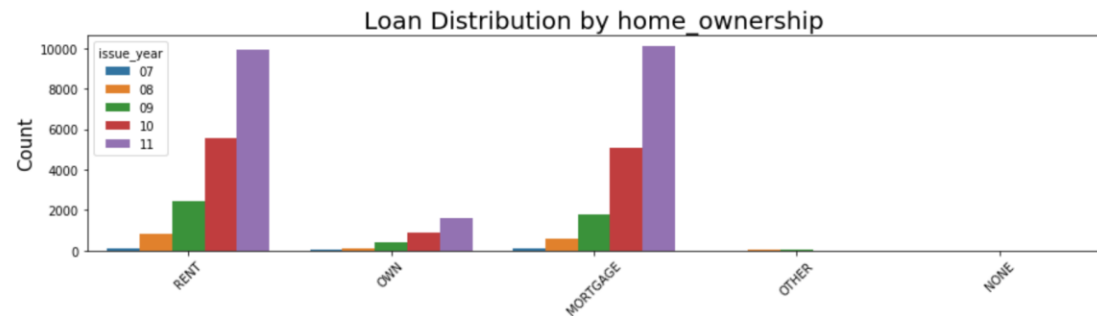
- Consistently the loan requests have been for debt management



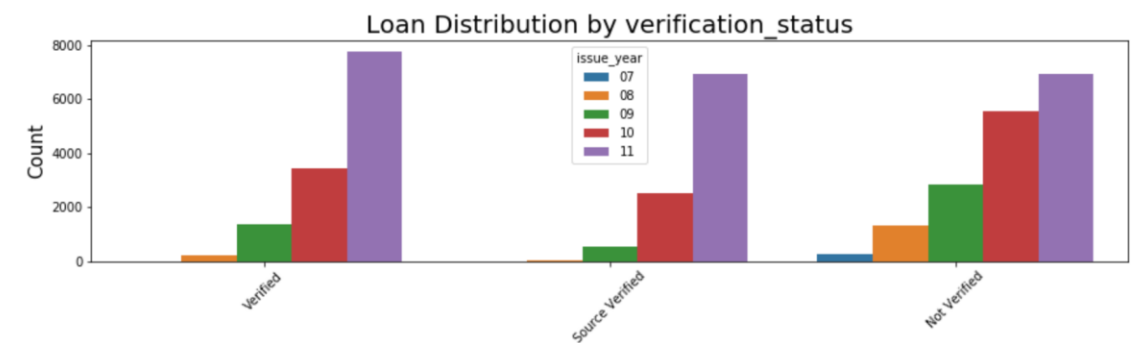
- Higher and consistent (over years) loan volumes from CA, NY, FL and TX



- Applicants who live in rental homes or with a mortgage are key loan seekers.

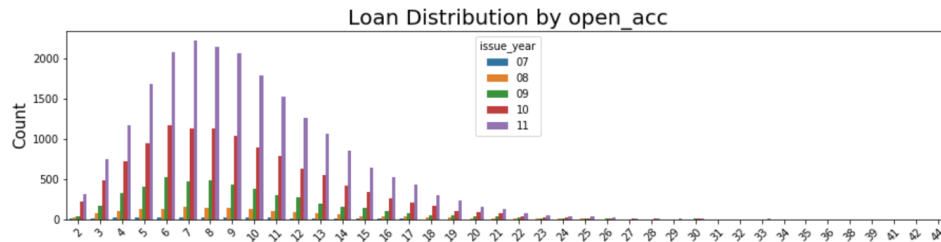


- Shows a trend of increasing “non-verified” loan approvals, increasing risk factor for the bank

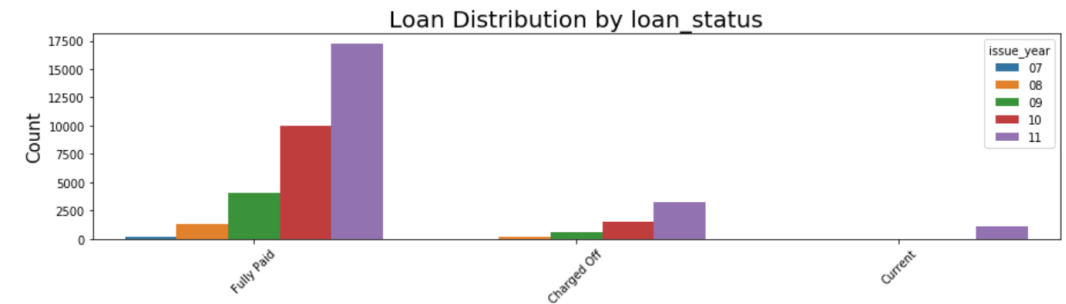


Univariate Analysis

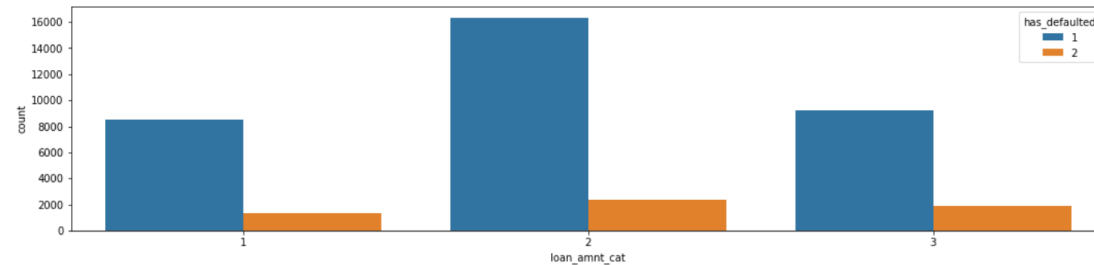
- People with 6 to 8 accounts are given most of the loan



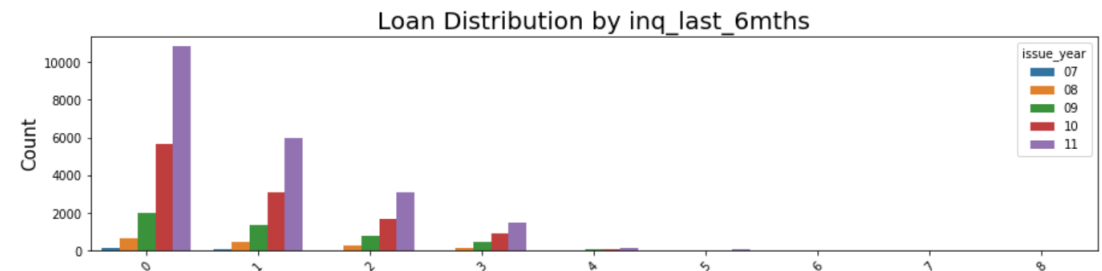
- 2011 is seeing higher default rate compared to previous years



- Loans in the range of 5500 to 15000 see higher default rate

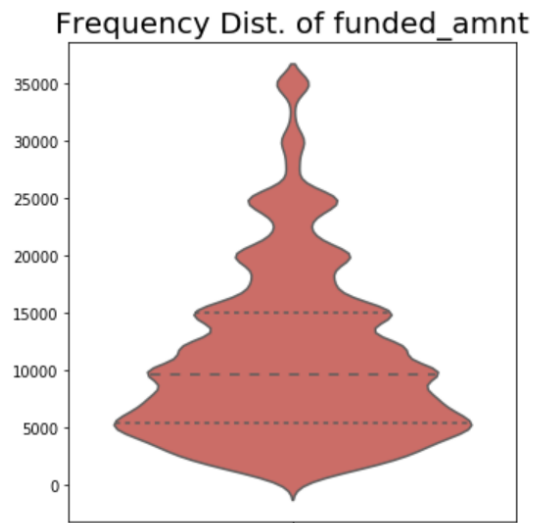


- Max loans given to people who didn't apply in last 6 month



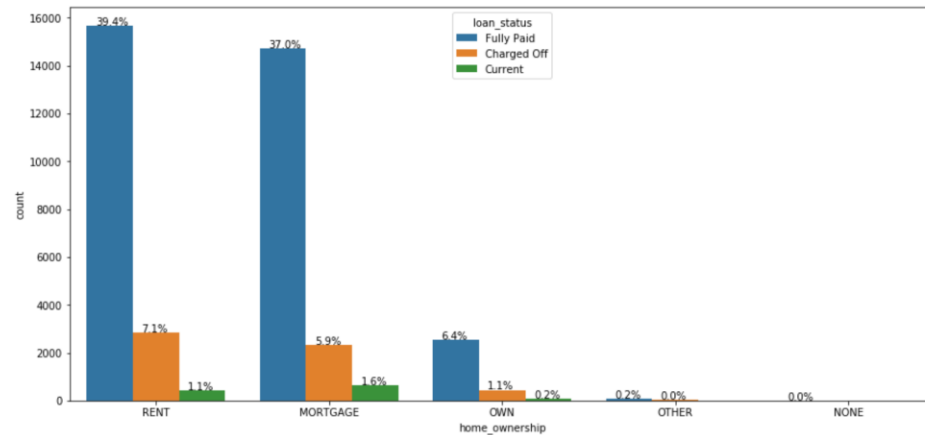
Univariate Analysis

- Maximum loans fall in the 5000 bucket

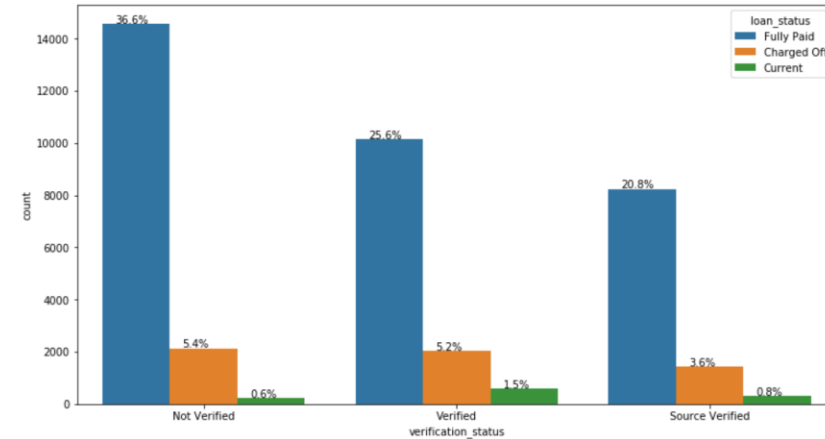


Bivariate analysis

- Loan applicants in rental home show higher repayment and default



- Default rate in verified loans are trending higher compared to



- Across the grades, loan for debt management stands out

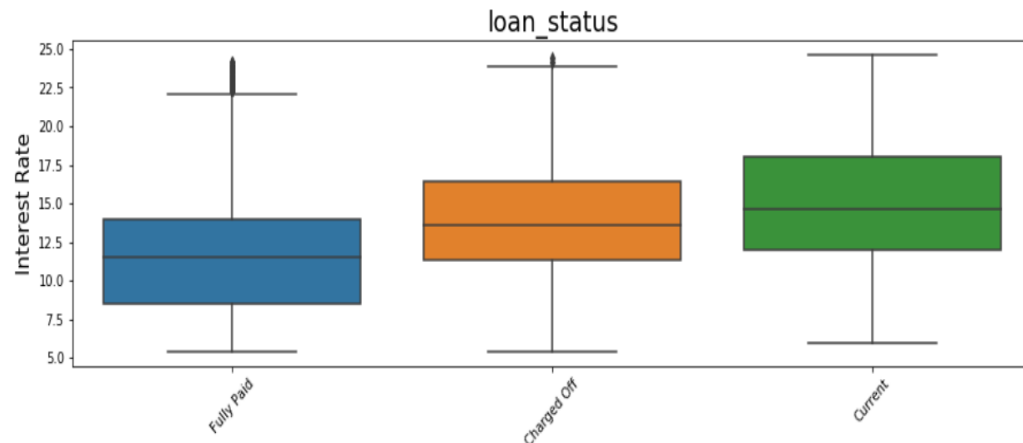
grade	A	B	C	D	E	F	G
purpose							
car	608	498	244	130	56	11	2
credit_card	1401	1652	1071	591	284	102	29
debt_consolidation	3871	5630	3954	2831	1589	596	170
educational	81	102	92	31	14	4	1
home_improvement	963	903	537	305	189	63	16
house	101	108	70	51	30	15	6
major_purchase	833	604	384	233	99	28	6
medical	210	204	136	84	42	14	3
moving	171	185	117	68	28	9	5
other	1089	1212	854	505	227	84	22
renewable_energy	30	37	17	9	5	3	2
small_business	348	476	355	301	204	97	47
vacation	134	126	66	31	21	2	1
wedding	245	283	201	137	54	21	6

Bivariate analysis

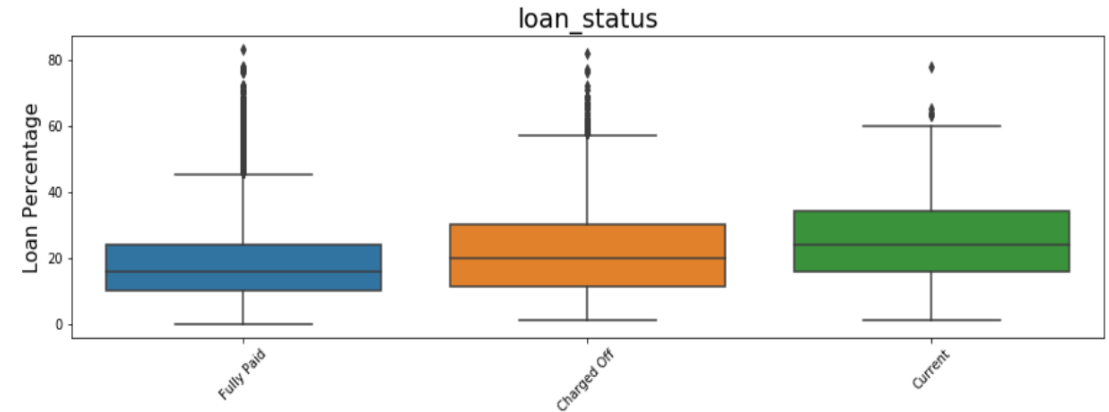
- Not verified applicants are more likely to default

loan_status	Charged Off	Current	Fully Paid
verification_status			
Not Verified	2142	227	14552
Source Verified	1434	310	8243
Verified	2051	603	10155

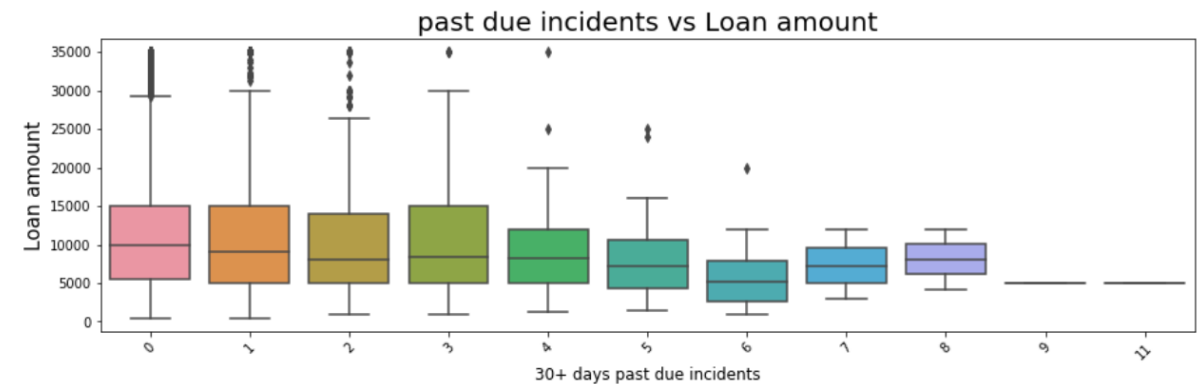
- Loan applicants with higher interest rate are more likely to be the defaulter



- Loan applicants with loan amount to annual income has higher percentage are more likely to default

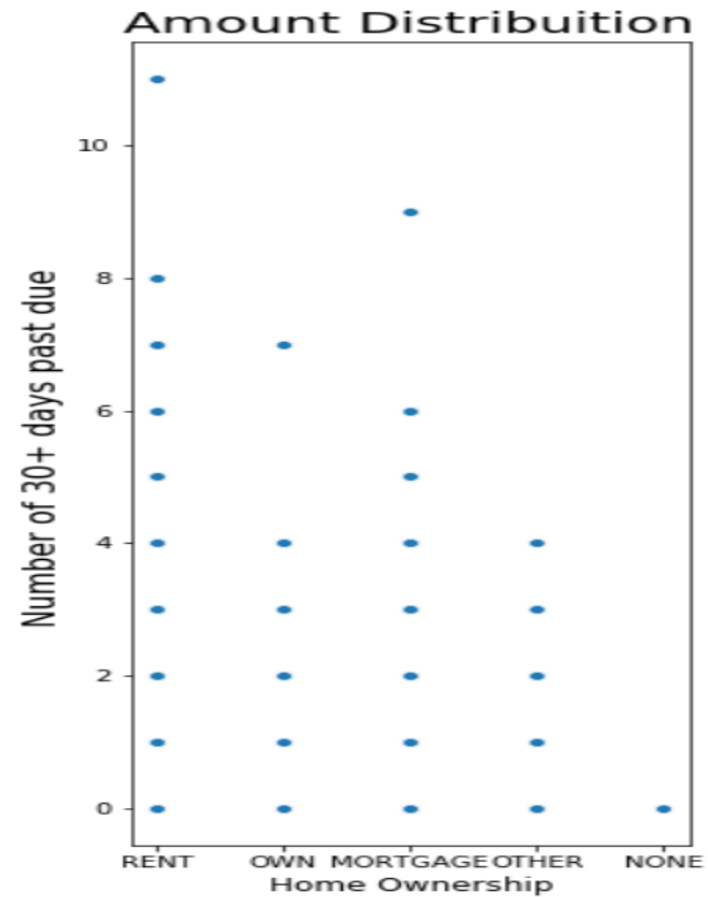


- Loan applicants with past due incidents more than 2 are more likely to default their loan amount



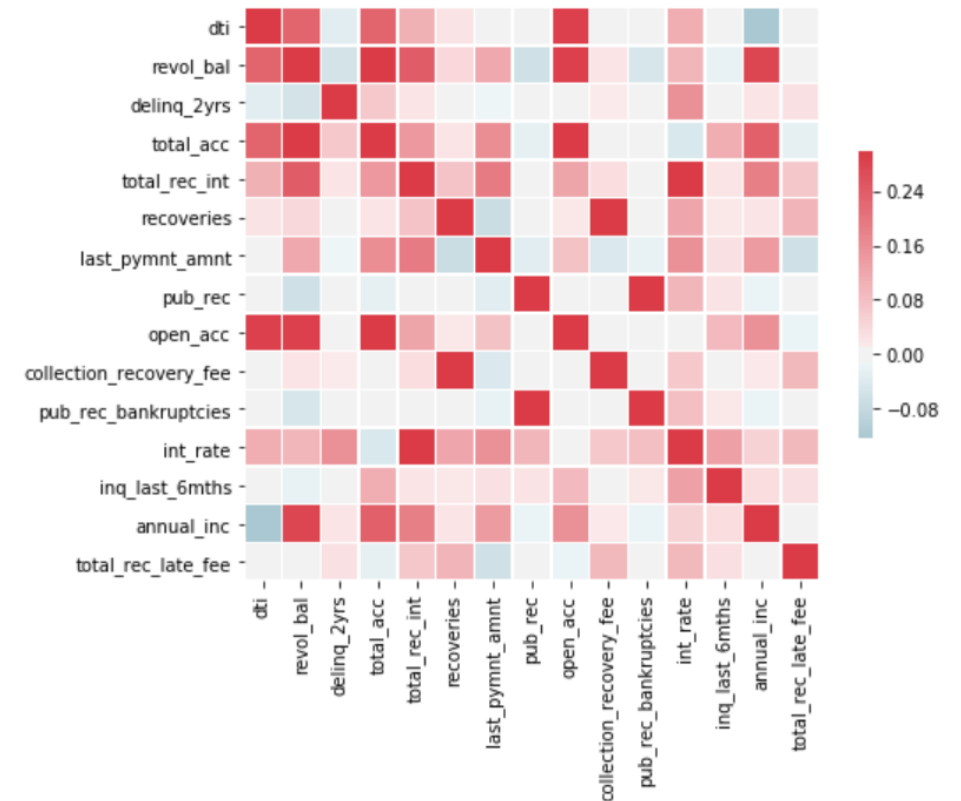
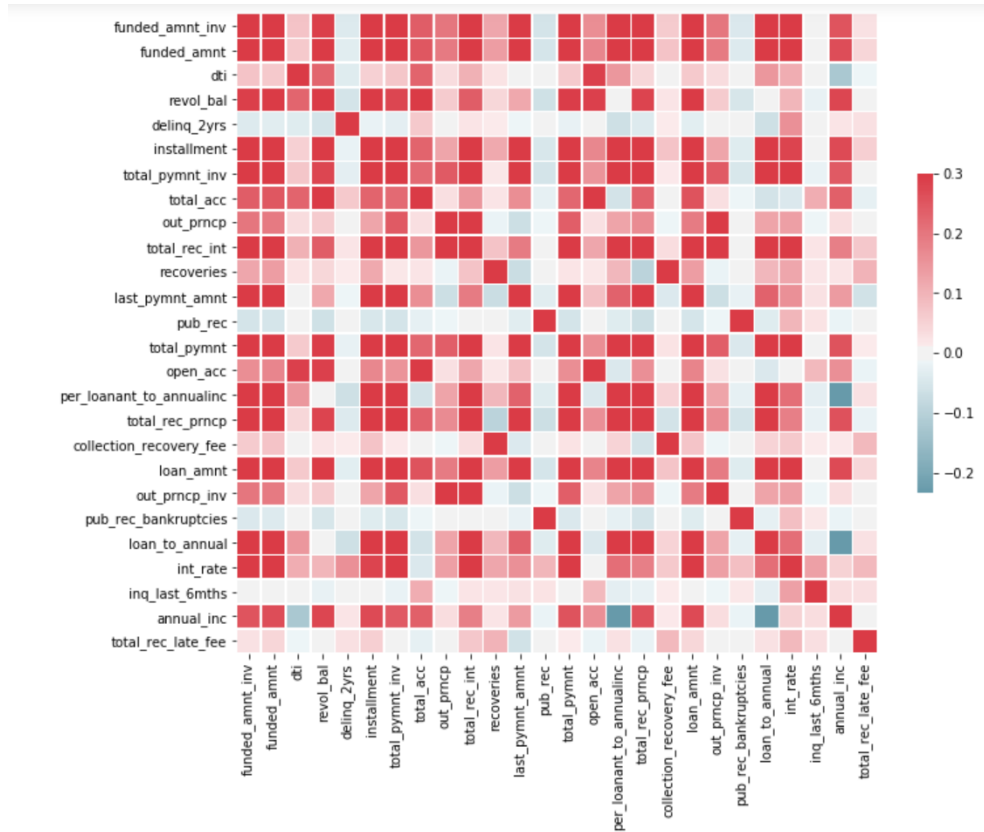
Bivariate analysis

- Loan applicants are with rented house has more number of past due EMI and are more likely to default the loan



Correlations Analysis

loan_amount is highly correlated with funded_amnt, installment, total_pymnt, loan_to_annual which actually true in real life case also recoveries is highly correlated with collection_recovery_fee total_rec_pncp is also correlated with total_pymnt_inv i.e principle is dependent on the investor amount



Summary

- Highest percent of defaulter was in 2011
- From 2010 onward 60 month term loan was provided
- Since 2008 count of non verified loan decreased marginally
- Maximum loans were provided to rent/mortgage clients
- In general higher grade clients were given more loans
- People with more that 10+ exp were provided highest no of loans
- Debt management was the main reason for taking loans
- Most amount of loan was issued at CA, NY, FL and TX states
- Max loans given to people who didn't apply in last 6 month
- People having 6-10 open accounts were given most no of loans
- People having more than 3 past due incidents are more likely to default their loan
- People with percentage of loan amount is more than 40 percent of annual income are more likely to default their loan