Lending Club Case Study

SUBMISSION

@Author:Baratchandar Venkatapathy and Barun Kumar Mishra

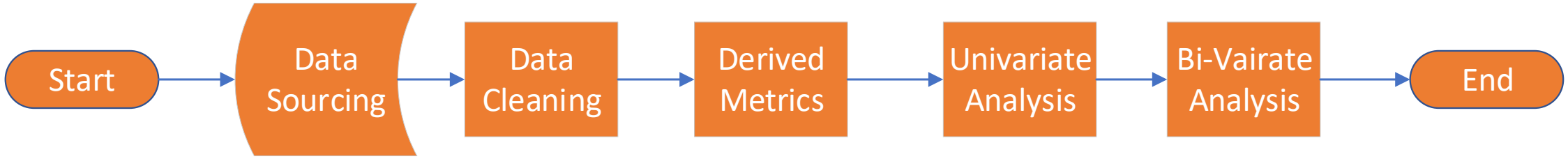
Business Objectives

- Lending Club is the largest online loan marketplace, facilitating personal loans, business loans and financing of medical procedures. Borrowers can easily access lower interest rate loans through a fast online interface.
- Lending Loan to Risky applicants is the largest source of financial loss(called credit loss). The credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed. In other words, borrowers who default cause the largest amount of loss to the lenders. In this case the customers labelled as “Charged-Off” are the “Defaulters”.
- The company wants to understand the driving factors(or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.

Data Understanding/Metadata

Meta Data	Exit Data				
Description	Loan Data for all loans issued through the time period 2007 to 2011				
Source	loan.csv				
Format	csv				
Number of rows	39717				
Each row is	Loan Data				
Sampling Method	Loan Data for all loans issued through the time period 2007 to 2011				
Prepared Data	23rd March				
We have done EDA Base on Below Columns					
Column Name	Type	Description	Missing	Uniques	Top/Mean
addr_state	Object	The state provided by the borrower in the loan application	NA	39717	CA
annual_inc	float64	The self-reported annual income provided by the borrower during registration.	NA	39717	6.90E+04
application_type	Object	Indicates whether the loan is an individual application or a joint application with two co-borrowers	NA	39717	INDIVIDUAL
collection_recovery_fee	float64	post charge off collection fee	NA	39717	12.4
dti	float64	A ratio calculated using the borrower's total monthly debt payments on the total debt obligations, excluding mortgage and the requested LC loan, divided by the borrower's self-reported monthly income.	NA	39717	13.31
earliest_cr_line	float64	The month the borrower's earliest reported credit line was opened	NA	39717	Nov-98
emp_length	float64	Employment length in years. Possible values are between 0 and 10 where 0 means less than one year and 10 means ten or more years.	NA	38642	10+ years
funded_amnt	float64	The total amount committed to that loan at that point in time.	NA	39717	10947
funded_amnt_inv	float64	The total amount committed by investors for that loan at that point in time.	NA	39717	10347.44
grade	Object	LC assigned loan grade	NA	39717	B
home_ownership	Object	The home ownership status provided by the borrower during registration. Our values are: RENT, OWN, MORTGAGE, OTHER.	NA	39717	RENT
installment	float64	The monthly payment owed by the borrower if the loan originates.	NA	39717	324.56
int_rate	float64	Interest Rate on the loan	NA	39717	10.99
issue_d	Object	The month which the loan was funded	NA	39717	11-Dec
last_pymnt_amnt	float64	Last total payment amount received	NA	39717	2678.82
last_pymnt_d	float64	Last month payment was received	NA	39646	16-May
loan_amnt	float64	The listed amount of the loan applied for by the borrower. If at some point in time, the credit department reduces the loan amount, then it will be reflected in this value.	NA	39717	11219.44
loan_status	Object	Current status of the loan	NA	39717	Fully Paid

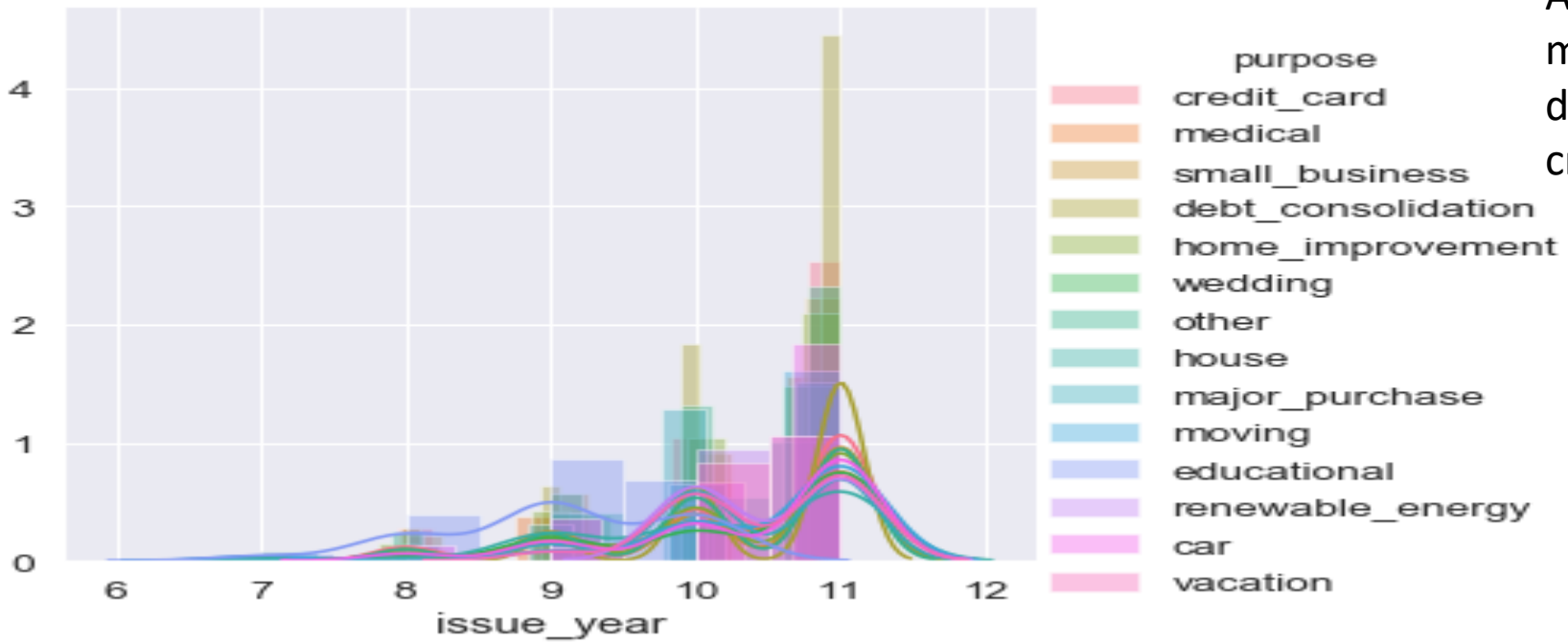
Data Analysis Flowchart



Univariate Analysis

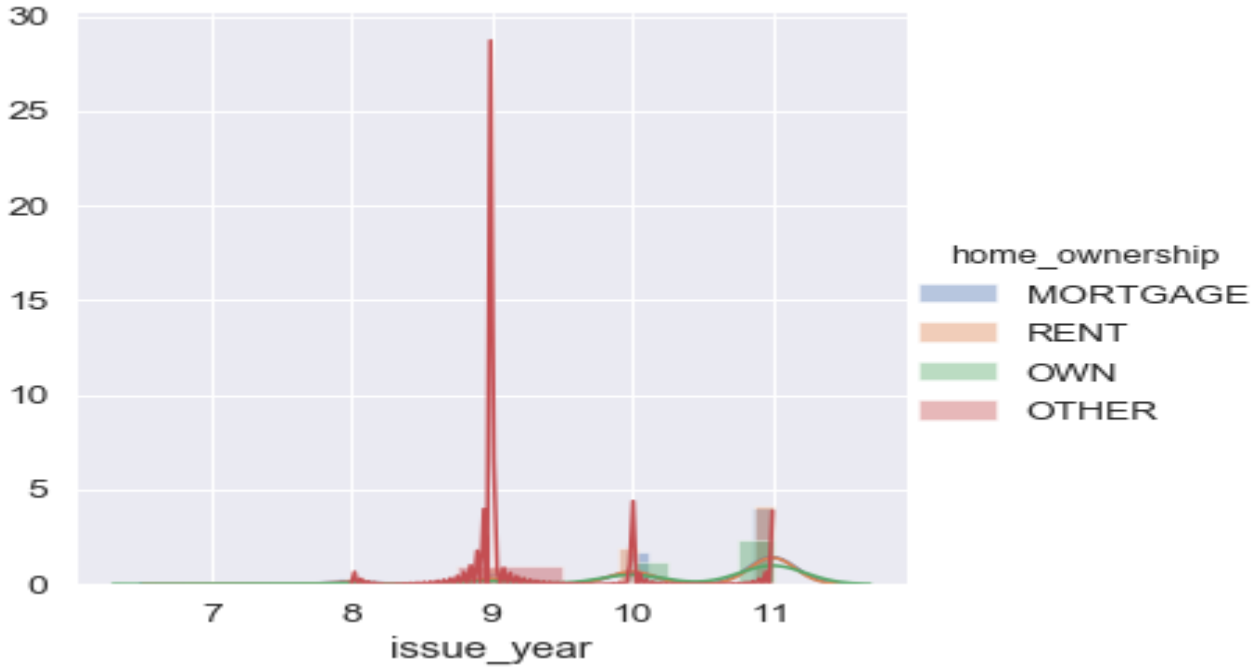
Loan Applicants have been increasing exponentially from year 2007 to 2011

Univariate Analysis



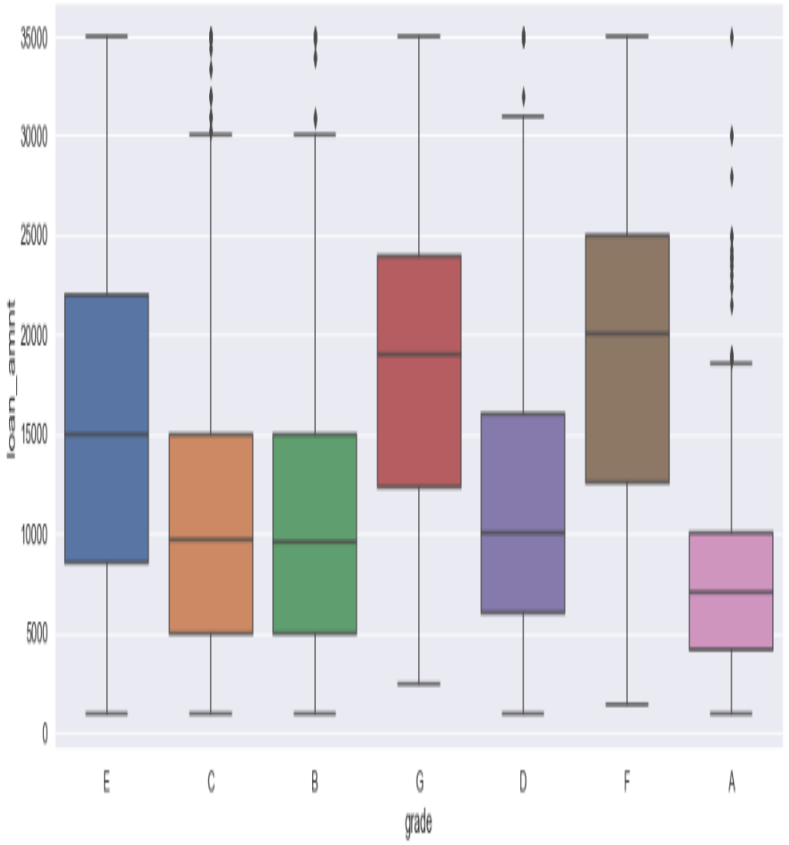
As per the chart, customer defaulted mostly who borrowed the purpose for debt_consolidation and followed by credit card purpose.

Segmented Univariate Analysis



As per the chart, based on the homeownership category segmented the defaulted customers loan data..
In each year the values defaulted are based on the customer's home ownership category

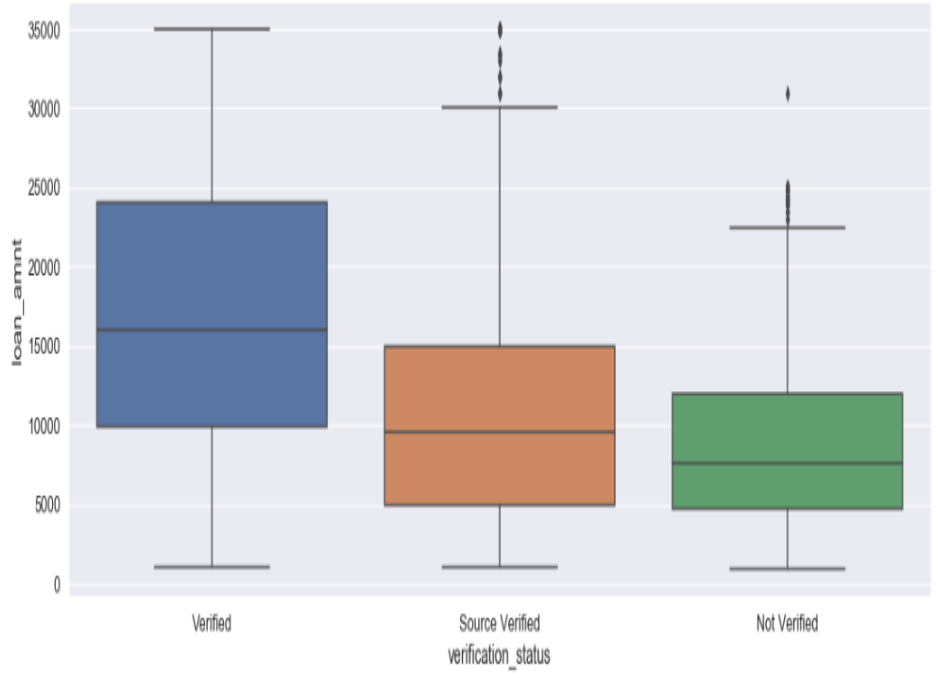
Bivariate Analysis



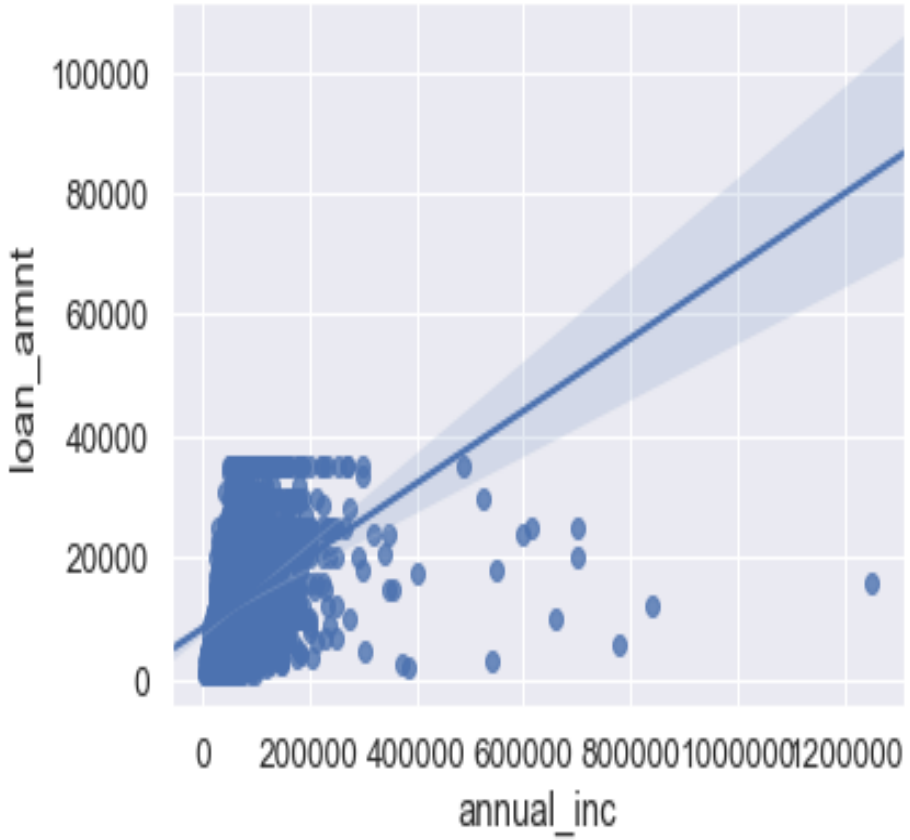
Box plot
Based on the customer grade vs loan amount

Customers who are graded B defaulted close to 25 % .
Customers who are graded C and D defaulted around 20 % .

Box plot
The source category “Source Verified” customers borrowed loan amount in the range of 5000 to 15000 are the top defaulted customers as compared to other source categories.



Bivariate Analysis



Scatted plot
Based on the Annual income and loan amount ,plotted this one

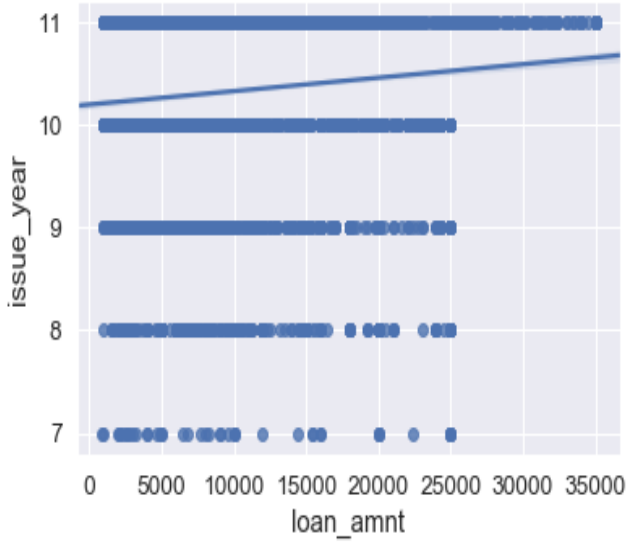
Here shows almost whose annual income range between 10000 to 20000 are defaulted

.

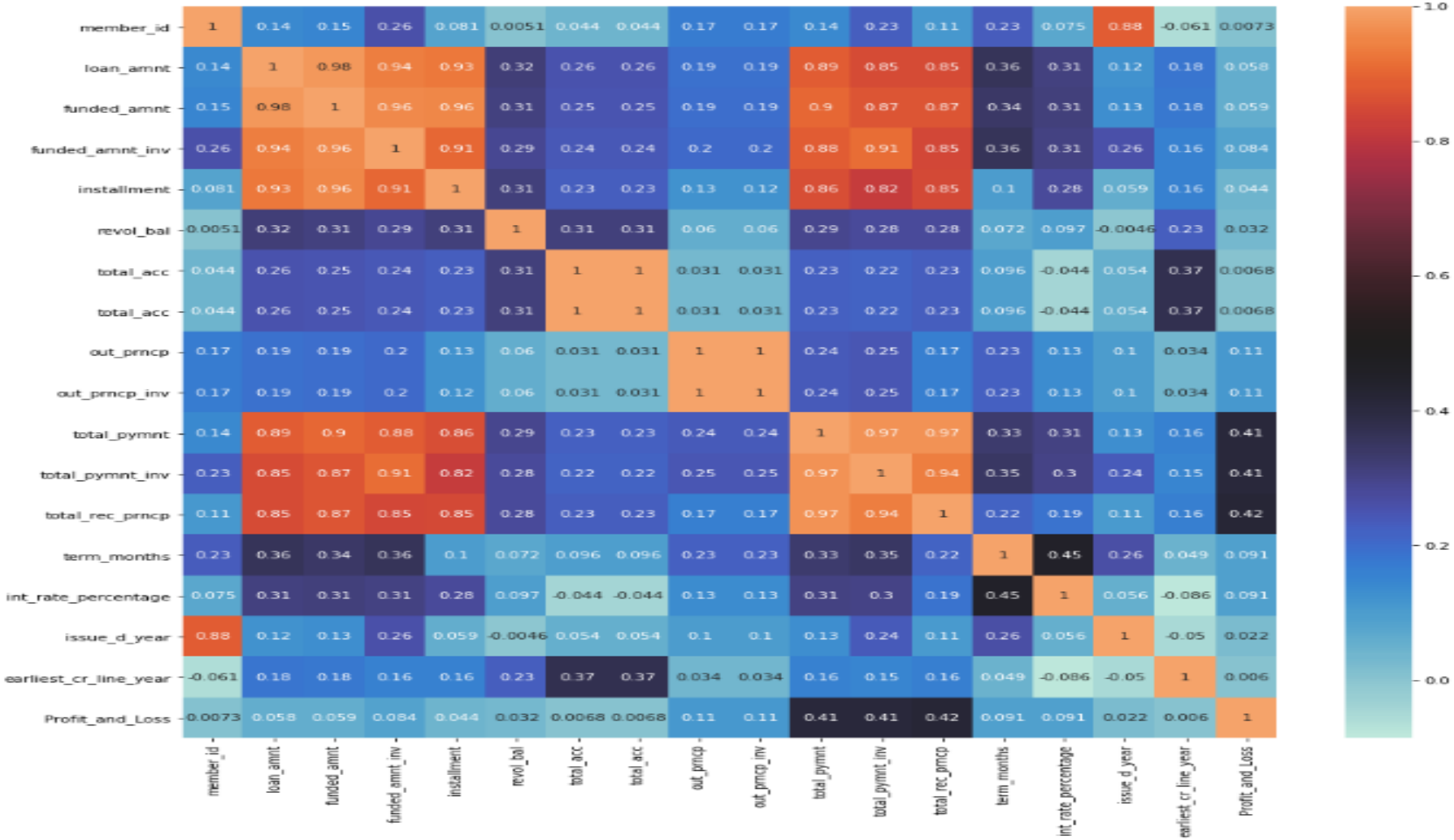
Scatted plot

Issue year vs Loan amount

Looks like most of the customers who borrowed in the year 2011 were defaulted.



Heat Map


member_id 1 0.14 0.15 0.26 0.081 0.0051 0.044 0.044 0.17 0.17 0.14 0.23 0.11 0.23 0.075 0.88 -0.061 0.0073
loan_amnt 0.14 1 0.98 0.94 0.93 0.32 0.26 0.26 0.19 0.19 0.89 0.85 0.85 0.36 0.31 0.12 0.18 0.058
funded_amnt 0.15 0.98 1 0.96 0.96 0.31 0.25 0.25 0.19 0.19 0.9 0.87 0.87 0.34 0.31 0.13 0.18 0.059
funded_amnt_inv 0.26 0.94 0.96 1 0.91 0.29 0.24 0.24 0.2 0.2 0.88 0.91 0.85 0.36 0.31 0.26 0.16 0.084
installment 0.081 0.93 0.96 0.91 1 0.31 0.23 0.23 0.13 0.12 0.86 0.82 0.85 0.1 0.28 0.059 0.16 0.044
revol_bal 0.0051 0.32 0.31 0.29 0.31 1 0.31 0.31 0.06 0.06 0.29 0.28 0.28 0.072 0.097 -0.0046 0.23 0.032
total_acc 0.044 0.26 0.25 0.24 0.23 0.31 1 1 0.031 0.031 0.23 0.22 0.23 0.096 -0.044 0.054 0.37 0.0068
total_acc 0.044 0.26 0.25 0.24 0.23 0.31 1 1 0.031 0.031 0.23 0.22 0.23 0.096 -0.044 0.054 0.37 0.0068
out_prncp 0.17 0.19 0.19 0.2 0.13 0.06 0.031 0.031 1 1 0.24 0.25 0.17 0.23 0.13 0.1 0.034 0.11
out_prncp_inv 0.17 0.19 0.19 0.2 0.12 0.06 0.031 0.031 1 1 0.24 0.25 0.17 0.23 0.13 0.1 0.034 0.11
total_pymnt 0.14 0.89 0.9 0.88 0.86 0.29 0.23 0.23 0.24 0.24 1 0.97 0.97 0.33 0.31 0.13 0.16 0.41
total_pymnt_inv 0.23 0.85 0.87 0.91 0.82 0.28 0.22 0.22 0.25 0.25 0.97 1 0.94 0.35 0.3 0.24 0.15 0.41
total_rec_prncp 0.11 0.85 0.87 0.85 0.85 0.28 0.23 0.23 0.17 0.17 0.97 0.94 1 0.22 0.19 0.11 0.16 0.42
term_months 0.23 0.36 0.34 0.36 0.1 0.072 0.096 0.096 0.23 0.23 0.33 0.35 0.22 1 0.45 0.26 0.049 0.091
int_rate_percentage 0.075 0.31 0.31 0.31 0.28 0.097 -0.044 -0.044 0.13 0.13 0.31 0.3 0.19 0.45 1 0.056 -0.086 0.091
issue_d_year 0.88 0.12 0.13 0.26 0.059 -0.0046 0.054 0.054 0.1 0.1 0.13 0.24 0.11 0.26 0.056 1 -0.05 0.022
earliest_cr_line_year -0.061 0.18 0.18 0.16 0.16 0.23 0.37 0.37 0.034 0.034 0.16 0.15 0.16 0.049 -0.086 -0.05 1 0.006
Profit_and_Loss 0.0073 0.058 0.059 0.084 0.044 0.032 0.0068 0.0068 0.11 0.11 0.41 0.41 0.42 0.091 0.091 0.022 0.006 1
member_id
loan_amnt
funded_amnt
funded_amnt_inv
installment
revol_bal
total_acc
total_acc
out_prncp
out_prncp_inv
total_pymnt
total_pymnt_inv
total_rec_prncp
term_months
int_rate_percentage
issue_d_year
earliest_cr_line_year
Profit_and_Loss
1.0
0.8
0.6
0.4
0.2
0.0

Conclusion

- Customer borrowed for the purpose “debt_consolidation” are almost 50 % defaulted. So organization need to reevaluate purpose type before initiating the loan.
- 25 % of Customers with Grade level “ B” are defaulted. Either need to reduce the loan amount for the customers who falls in Grade “B” or not to release the funds.
- Most of them whose annual income less than 200000 are defaulted when compared to others with annual income greater than 200000. Also loan amount value is around 10000 to 30000. Considering annual income may be need to redefine the business logic behind loan percentage amount issued based on the annual income.

  ex: Approve 8 to 12 % of total income as loan amount vs Approve 10 to 25 % of total income as loan amount

Suggestion here is to reduce the loan amount based on the annual income

- Looks like most of the customers from the year 2011 were defaulted. So there is crisis or something external happened outside . Organization need to reevaluate the loan policy periodically considering external factors like Unemployment ,Job market, crisis