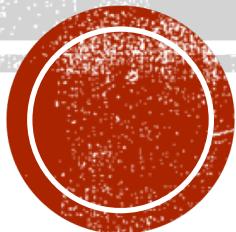


LENDING CLUB ASSIGNMENT SUBMISSION

Aman Singh Bains



PROBLEM STATEMENT

- For our consumer finance company, we need to analyze existing data of past and ongoing loans so that we can devise a **Risk Analysis framework** to facilitate:-
 - Avoiding to give out Loans where probability of default is higher
 - Providing Loans where probability of default is lower
- Thus, maximizing growth and lowering risk significantly



SOLVING APPROACH

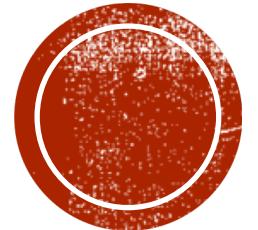
- Firstly Cleanup the data to enable analysis
- Create meaningful calculated attributes to help with analysis
- Find patterns where default is higher and lower
 - Analyse by Categorical attributes
 - Analyse Spread of Numerical attributes
 - Analyse correlation of numerical attributes to defaults
- Identify attributes that effect defaults. Some attributes are found before giving loan and some are found during loan lifecycle
- Create a framework for Risk Analysis so that lending Company can
 - Avoid bad loans
 - Encourage good loans
- Monitor ongoing Loans and raise alerts if risk increases to mitigate risk



SOLUTION

- Create a **Customer Scoring** mechanism to calculate the score of each customer before deciding on giving the loan or not
 - Higher score would mean lower risk
 - Higher scores would lead to more Approvals
 - Lower scores would lead to more Rejections
- We will also monitor ongoing Loans and the customer during the lifecycle of the loan to **Raise Alerts** for the Company if a default is going to occur
 - For this we will setup alerts for specific attributes occurring





CUSTOMER SCORING

Analysis for Customer Scoring for Underwriting

SCORING - HOME OWNERSHIP

home_ownership	defaulted	count	default_percent
OTHER	18	98	18.37
RENT	2839	18899	15.02
OWN	443	3058	14.49
MORTGAGE	2327	17659	13.18
NONE	0	3	0.00

- We can give lower score for higher default percentage
- Other are worst & Mortgage are best
- Home Ownership Scores
- Other = 0
- Rent = 1
- Own = 2
- Mortgage = 3



SCORING - PURPOSE

	defaulted	count	default_percent
purpose			
small_business	475	1828	25.98
renewable_energy	19	103	18.45
educational	56	325	17.23
other	633	3993	15.85
moving	92	583	15.78
house	59	381	15.49
medical	106	693	15.30
debt_consolidation	2767	18641	14.84
vacation	53	381	13.91
home_improvement	347	2976	11.66
credit_card	542	5130	10.57
car	160	1549	10.33
major_purchase	222	2187	10.15
wedding	96	947	10.14

- We can give lower score for higher default percentage
- Small Business is worst and best are credit card, car, major purchase & wedding
- Scoring Purpose
 - Small Business = 0
 - Renewable Energy = 1
 - Educational = 2
 - ... so on



SCORING – VERIFICATION STATUS

verification_status	defaulted	count	default_percent
Verified	2051	12809	16.01
Source Verified	1434	9987	14.36
Not Verified	2142	16921	12.66

- We can give lower score for higher default percentage
- Verified are worst and Not verified are best
- Scoring Verification Status
 - Verified = 0
 - Source Verified = 1
 - Not Verified = 2
- Actually it looks like staff felt the need to verify because the loan seemed risky



SCORING - STATE

addr_state	defaulted	count	default_percent
NE	3	5	60.00
NV	108	497	21.73
AK	15	80	18.75
SD	12	64	18.75
FL	504	2866	17.59
ID	1	6	16.67
MO	114	686	16.62
HI	28	174	16.09
NM	30	189	15.87
CA	1125	7099	15.85
OR	71	451	15.74
UT	40	258	15.50
MD	162	1049	15.44
GA	215	1398	15.38
WA	127	840	15.12
NJ	278	1850	15.03
NH	25	171	14.62
NC	114	788	14.47
MI	103	720	14.31

- We can give lower score for higher default percentage
- States are listed worst to best
- Scoring State
 - We can start giving score for state from 0 for NE and increasing by 1 for each next state with lower default percentage
 - But keeping the score same for same percentage default ignoring the decimal points



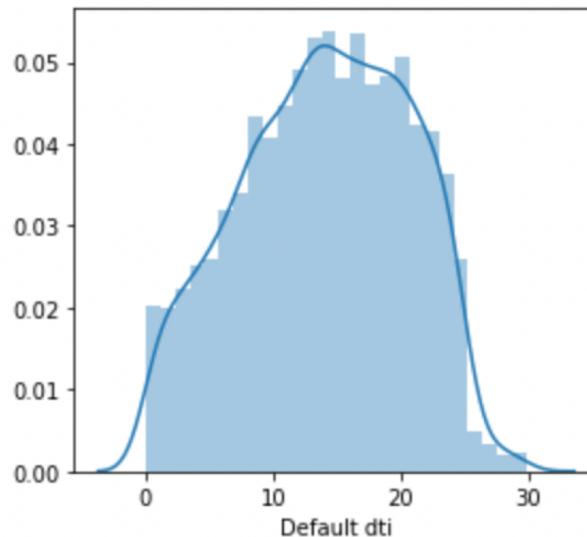
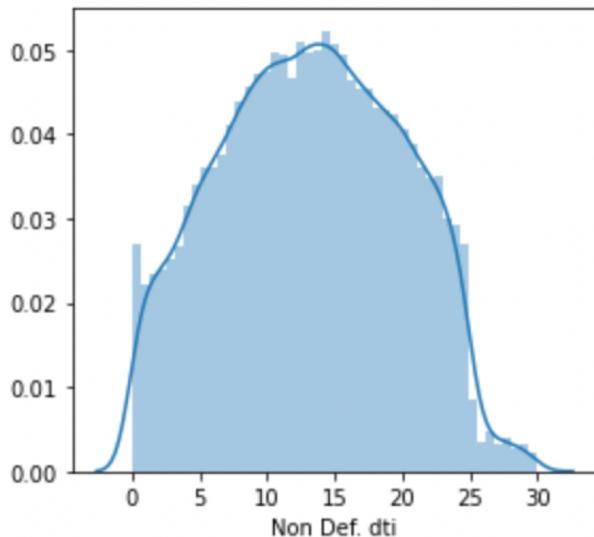
SCORING – ZIP CODE

addr_state	zip_code	defaulted	count	default_percent
OK	746xx	5	7	71.43
MN	561xx	5	7	71.43
IL	607xx	7	14	50.00
	608xx	4	8	50.00
AR	719xx	5	11	45.45
CO	808xx	5	11	45.45
OK	744xx	5	11	45.45
MI	496xx	3	7	42.86
NV	897xx	3	7	42.86
KY	425xx	2	5	40.00
SD	573xx	2	5	40.00
MI	499xx	2	5	40.00
OH	438xx	2	5	40.00
KY	416xx	2	5	40.00
MD	215xx	3	8	37.50
WV	253xx	4	11	36.36
IL	611xx	5	14	35.71
CA	924xx	7	20	35.00
GA	316xx	6	18	33.33
MO	639xx	2	6	33.33
CA	912xx	16	48	33.33
TX	755xx	2	6	33.33
WV	264xx	2	6	33.33
NM	883xx	3	9	33.33

- Zip codes with loan count higher than 5 & with highest to lowest default rate were extracted
- Scoring zip_code
- Company should assign lower scores for zip codes with higher default
- There are always good and bad localities, those with higher default are bad localities, loans from these localities should get lower scores so that defaults can be avoided



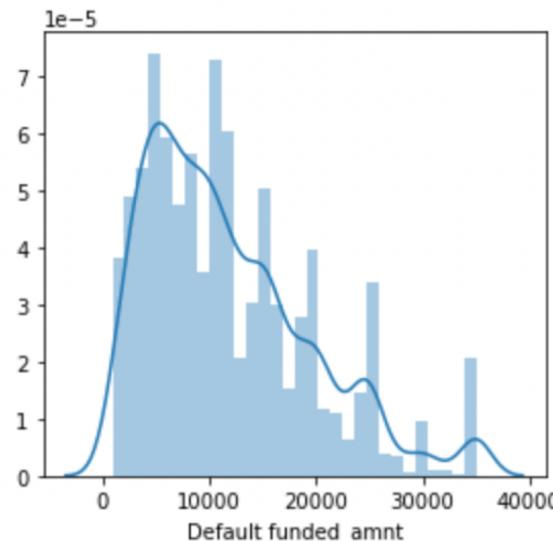
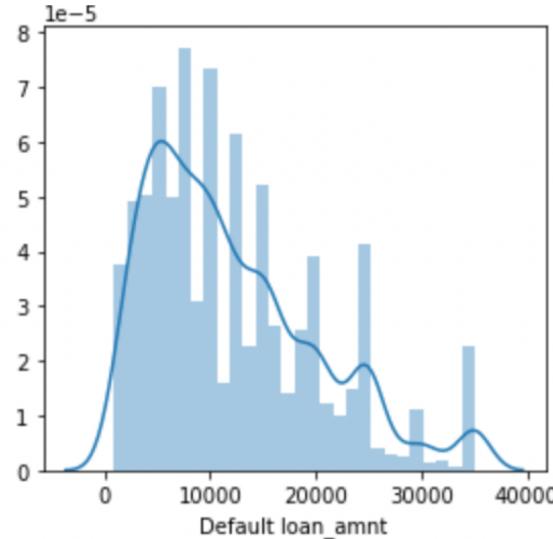
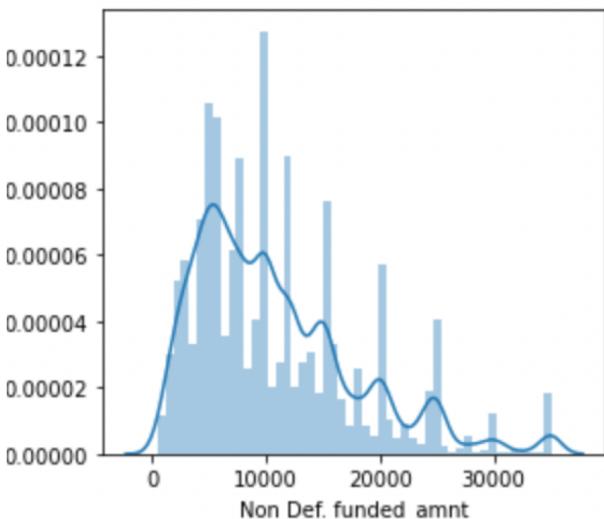
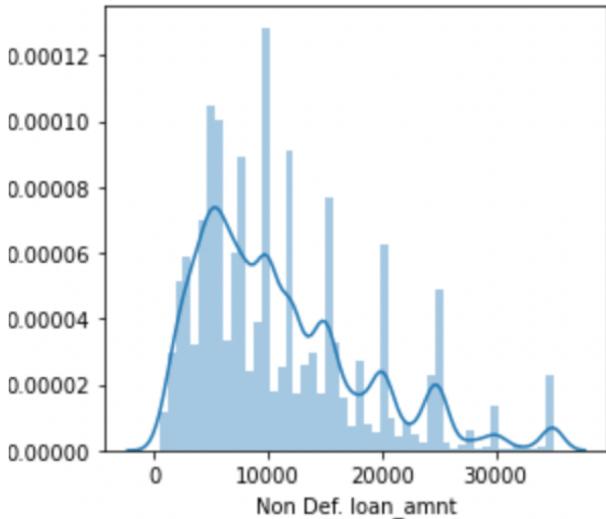
SCORING – DEBT TO INCOME RATIO



- Left chart is for Non Default while right chart is for Default
- Higher Debt to Income Ratio leads to more default
- If Debt is higher in comparison to Income then default is more
- Scoring dti
 - We can score dti as $100 - dti$
 - This will lead to higher score for lower dti



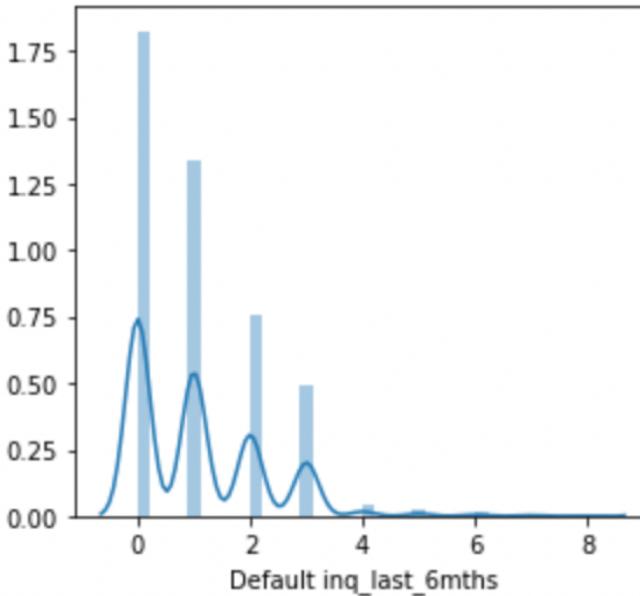
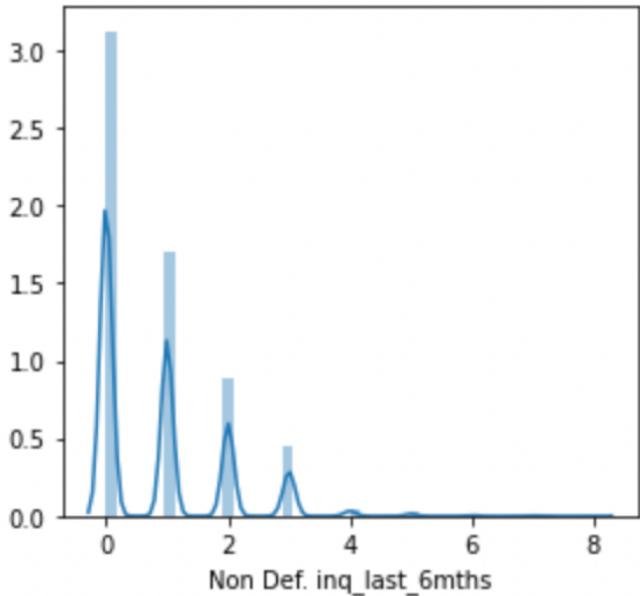
SCORING – LOAN AMOUNT



- Top charts are for Loan Amount, bottom charts are for Funded amount, left charts are for Non Default, right charts are for Defaulted Loans
- Loan Amount & Funded Amount is higher in case of Defaulted Loans as in chart on the right
- Scoring Loan Amount
 - As per each Purpose a safe range of loan amount needs to be defined
 - For this range score should be 1
 - Above that range score should be 0 to discourage
 - Below that range score should be 2 to encourage



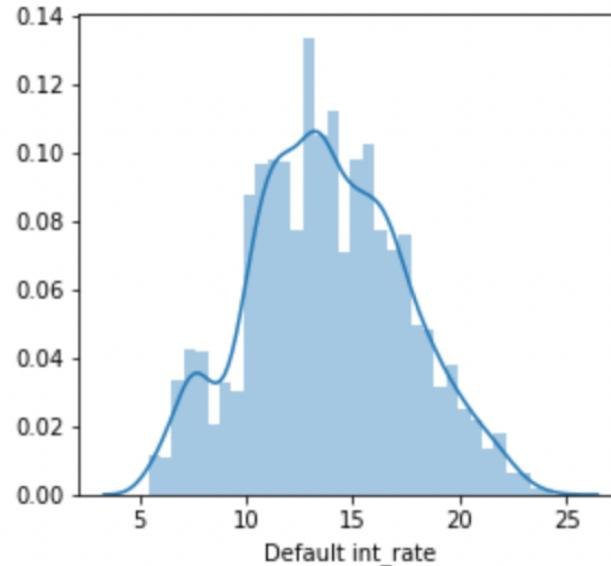
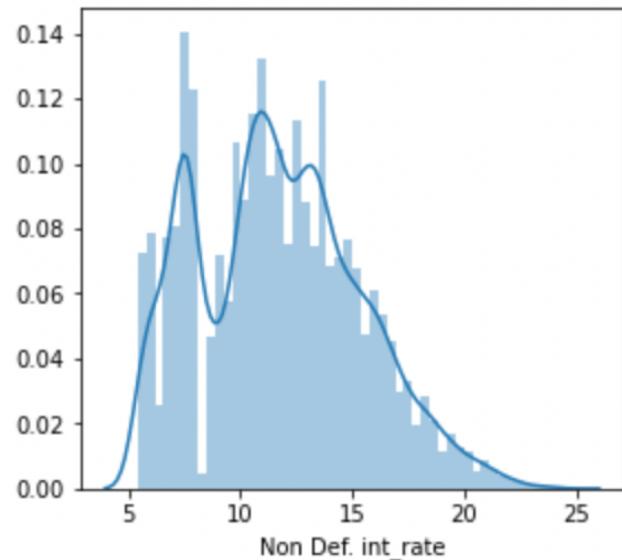
SCORING – INQUIRIES FOR LOANS



- Inquiries for Loans in last 6 months is higher for Defaulted Loans as per right chart
- It can be deduced that when a Customer is Inquiring from here and there for Loans that means that he is in financial trouble and Loan Default happens
- This information can be received from the credit bureau
- Scoring – Higher enquiries should have lower score.
- Score = 10 – inquiry count. As number would be less than 10



SCORING – INTEREST RATE



- Company is offering higher rate to risky Customers, this should also lead to lower score
- Chart on right shows higher interest rate for defaulted loans
- Scoring
- Score = highest rate – applicable rate
- This would discourage higher rate loans and lead to more rejections for higher rate if other factors are also negative



SCORING - BANKRUPTCIES

- Higher bankruptcies lead to more default
- Incase of a Customer has already declared Bankruptcy the Loan should be outrightly rejected

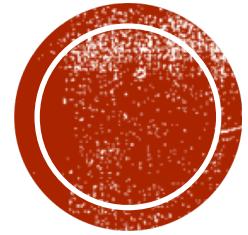
pub_rec_bankruptcies	defaulted	count	default_percent
2.0	2	7	28.57
1.0	366	1674	21.86
0.0	5141	37339	13.77



CUSTOMER SCORING - CONCLUSION

- At the time of underwriting customer score should be calculated by the software
- A threshold score is to be decided for automatic approvals above that score
- A range of score should be decided below auto approval score for human assisted underwriting, Credit Manager can consider subjective arguments in favor and against approving and decide. Higher loan amounts should be mostly rejected in this range
- Below the above stated lower range all loans should be automatically rejected by the software
- Once Customer Scoring is in place then the Company will have a solid framework for Risk Analysis for underwriting





ALERTS – LOAN MONITORING

Company needs to setup several alerts during loan lifecycle to take pre-emptive action in case of alerts coming in for any loans.

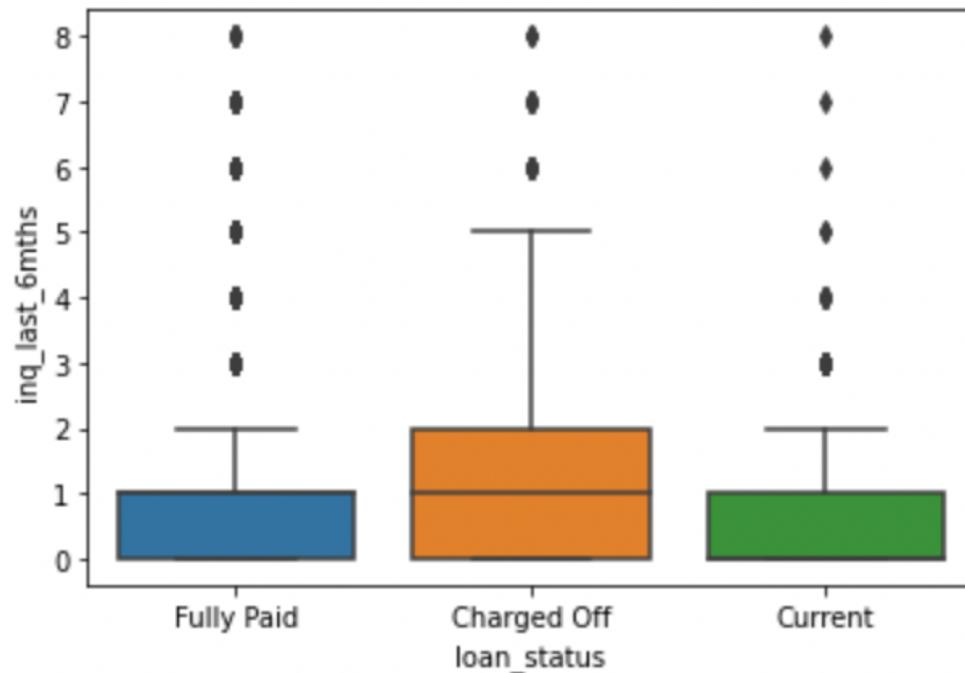
ALERT - BANKRUPTCY

pub_rec_bankruptcies	defaulted	count	default_percent
2.0	2	7	28.57
1.0	366	1674	21.86
0.0	5141	37339	13.77

- If customer declares bankruptcy then alert should be raised
- This information can be received from time to time from credit bureau
- This is the most **severe alert**
- Company should be alerted so that they can recover as much of the amount by repossessing asset financed or through legal means by filing a court case/ Arbitration



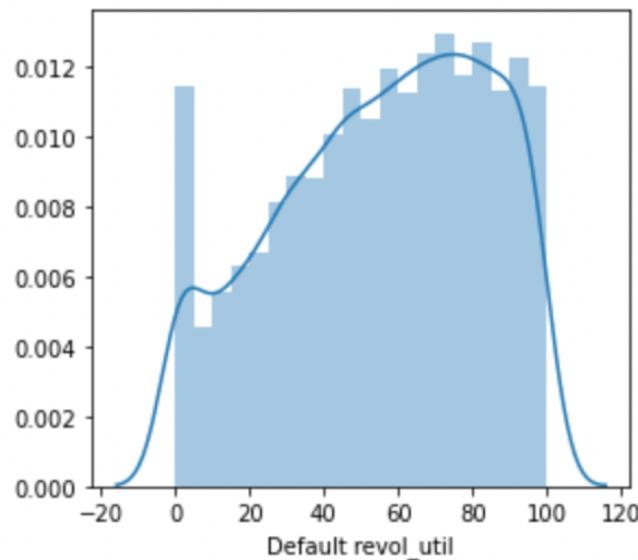
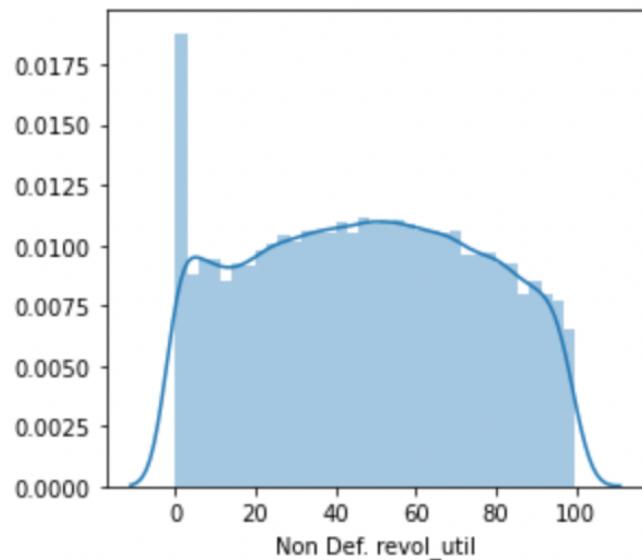
ALERT – INQUIRIES FOR LOANS



- We can receive this information from credit bureau from time to time
- In case of new Loan Inquiries by Customers who already have loans with the Company an alert should be raised
- Close monitoring is required in this case
- Should encourage the customer to repay the loan if possible through incentives



ALERT – REVOLVING UTILITY



- Right chart shows higher revolving utility for Defaulted Loans
- Company should monitor this attribute during loan lifecycle and raise alerts
- Close monitoring is required in this case
- Should encourage the customer to repay the loan if possible through incentives



ALERTS - CONCLUSION

- There can be several more alerts that can be setup during loan lifecycle like Delinquency in last 2 years, pub_rec
- Alerts will help in lowering the disaster just in time

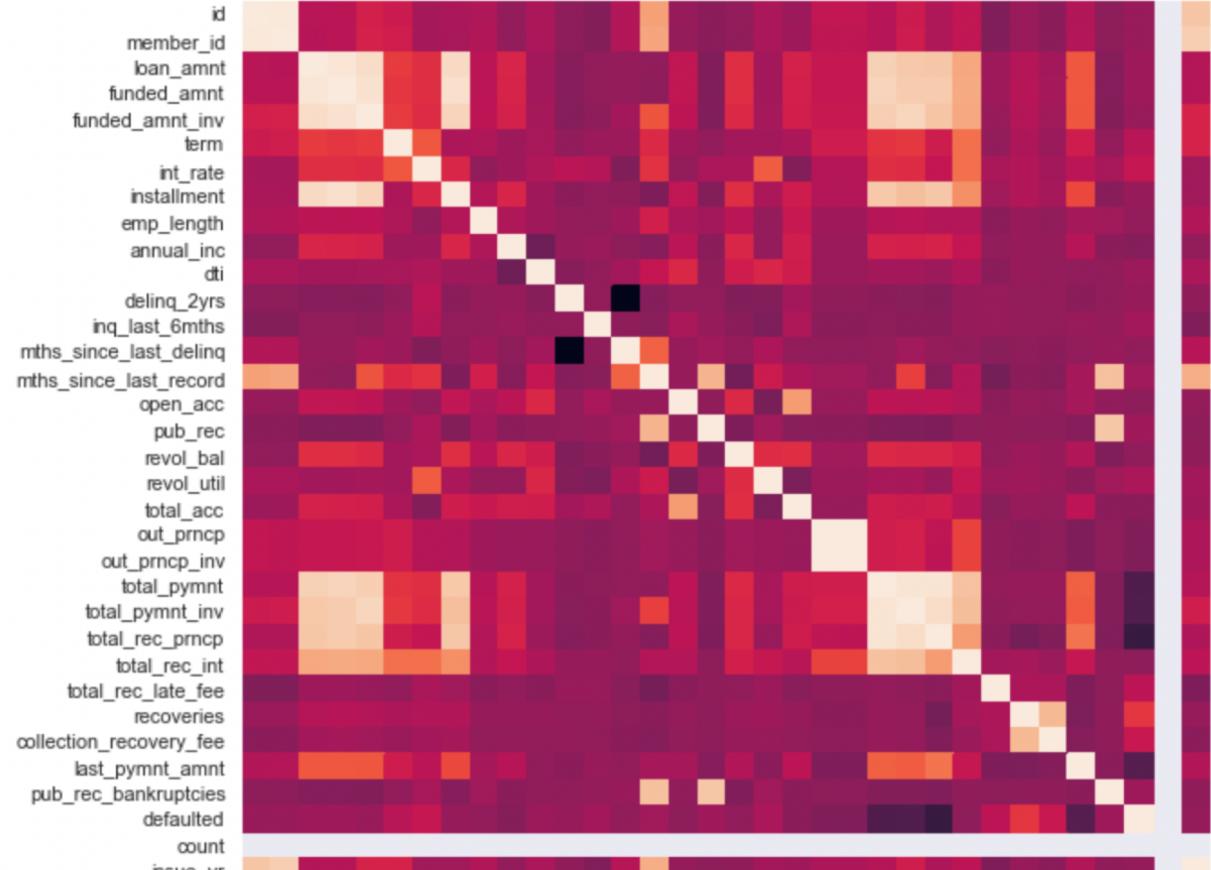


OBSERVATIONS

- Home ownership = Rent has highest occurrence, Mortgage comes second
- Not Verified loan count is highest, for Defaults Verified & Source verified is 2nd highest
- Purpose Debt Consolidation frequency is highest
- CA California has highest number of Loans
- In 2011 Company did highest number of loans
- Company Started operations in 2007 and there was a learning cost as default rate is highest in this year at approx. 18 %
- In 2008 & 2011 also Company had higher Default at 15% +



OBSERVATIONS



- **defauluted has some -ve correlation with annual_inc, total_rec_prncp, total_pymnt, last_pymnt_amnt**
 - **defauluted has correlation with recoveries, collection_recovery_fee, total_rec_late_fee & int_rate**

CONCLUSION

- The 3 Initiatives can lead to desired results
 - Customer Scoring before giving loans
 - Alerts during loan lifecycle
 - Charging late fee and collection charges in case of Default to discourage default. This has been observed in the data and should be made systematic for better effect

