BUSINESS OBJECTIVES

This company 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.

Like most other lending companies, lending loans to ‘risky’ applicants is the largest source of financial loss (called credit loss). 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'.

If one is able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. Identification of such applicants using EDA is the aim of this case study.

In other words, 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 utilize this knowledge for its portfolio and risk assessment.

Variable Understanding

❖ ***Variables are defined in Four border category***

- ❑ ID
- ❑ Internal
- ❑ Bureau
- ❑ Demographic

❑ **ID Variable**

A. ID :- Loan ID
B. Member ID :- Customer ID
C. Loan Status :- Status on the loan

❑ ***Internal***

A. last_credit_pull_d :- Bureau pull on which date.
B. loan_amnt :- Customer asked amount.
C. Funded_Amnt :- Total amount approved by lending club.
D. funded_amnt_inv :- Actual loan amount given to customer by lender.
E. grade :- grade or type of the loan.
F. sub_grade :- sub_grade of the loan.
G. term :- duration of the loan.
H. title :- loan title provide by the customer.(Unstructured format)
I. verification status :- income is verified or not.
J. installment :- emi of the loan.
K. int_rate :- interest rate on the loan.
L. issue_D :- issue date of the loan.

❑ *Bureau*

A. Delinq_2yrs :- number of times dpd 30 plus in last 2 years.
B. earliest_cr_line :- date of first loan taken.
C. inq_last_6_months :- inquiry in last 6 months.
D. mths_since_last_delinq :- how many month the customer is clean with out any delinquency.
E. mths_since_last_record :- what is last time customer is recorded in bureau.
F. open_Acc :- number of active loans of the customer.
G. pub_rec :- Number of derogatory public records
H. pub_rec_bankruptcies :- Number of public record bankruptcies
I. revol_bal :- Credit revolving amount. Loan taken of 50K when is reducing to match the same taken another loan to make amount back to 50K.
J. total_acc :- total number of trade lines a customer have.
K. dti :- 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.

❑ *Demographic*

A. addr_state :- state of the customer.
B. decs :- purpose of the loan describe by the customer. (Unstructured format)
C. home_ownership :- If customer have home or not. Not verified field.
D. Purpose :- drop down field of purpose of customer.
E. zip_code :- hashed zip code of the customer.
F. verfication_status :- Income is verified or not.

EVENT DISTRIBUTION

Event Distribution		
Event	**Count Of Customers**	**Event Percentage**
Fully Paid	32950	83.0%
Charged Off	5627	14.2%
Current	1140	2.9%

Event Distribution after cleaning		
Event	**Count Of Customers**	**Event Percentage**
Fully Paid	32915	85.4%
Charged Off	5627	14.6%

RESULTS OF SIMPLE UNIVARIATE AND BIVARIATE ANALYSIS

Univariate Analysis

1. According to IQR almost all the Continuous column have less than 10% of outliers.
2. Most of the customer belongs to CA State(18%).
3. Last credit pull is 2016 for 37% of the total base.
4. For customer coming to lending club almost 47% of the customer purpose of taking loan is debt consolidation.
5. Disbursement is increasing year on year from 2007 to 2011. In 2011 53% of total disbursement happens.
6. We have verified income for almost 57% of the total customers.
7. According to our customer almost 89% customer owns a property or have a mortgage loan on them.
8. Almost 22% of the total customer have a employment length of 10 or 10+ years.
9. A4, B3, A5 are the most sold subgrade of loan category.
10. B,A,C are the most sold grade of the loan.
11. 75% of the total customer takes short term loan of 36 months.
12. Median loan_amount ask by the customer is 9600.
13. Median funded_amount suggested by lending club is 9550.
14. Median amount funded by investor is 8733.
15. Average intereset rate charge by the company is 11-12%.
16. Average installment is 255 per month.
17. Median anual income of the customer base is 58000.
18. dti is normally distributed accross customer.
19. for most of the customer delinq_2yrs is 0.
20. almost 50% of the population does 0 inquiry in last 6 months.

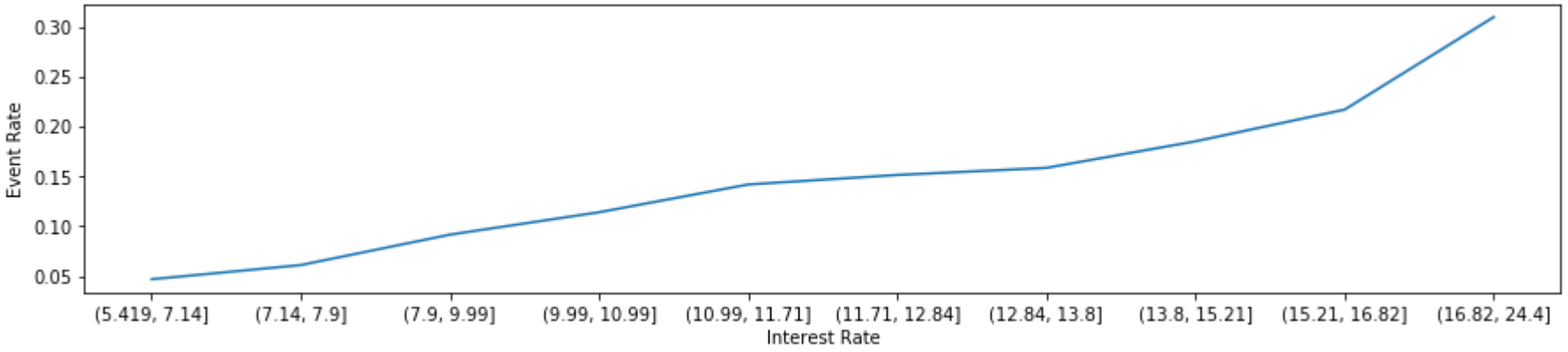
Bivariate Analysis

1. mths_since_last_delinq, open_acc, revol_bal are not related to event variable and they have very less p value.
2. term has the highest impact on loan status. but it has only two value and we cannot recommend bank to give only 36 or 60 months loan. (either one).
3. Higher the ask of loan_amnt higher chances of default same with funded amount and invested amount.
4. Higher term customers tends to default more.
5. Customer having higher int rate tends to default more. Higher int rate given to those customer who have high chance of default. This is moving in the correct direction.
6. G grade have the highest default rate.
7. People having missing emp length have more default rate.
8. Home ownership having value as other tends to default more.
9. Higher the income tends to default less.
10. Verified income salary people tends to default more. Reason can be non verified income employee come from small business background.
11. People taking loan for education tends to default more.
12. higher dti higher chances to default more.
13. Higher the 2 year delinquency higher chances of default.
14. High inquiry higher chances of default.
15. Higher number of open account lower the chance of default. Knows how to handle money. same with total number of loans.
16. Higher Number of public record bankruptcies higher chances of default.

SUGGESTIONS (AT THE TIME OF COLLECTION)

Interest Rate

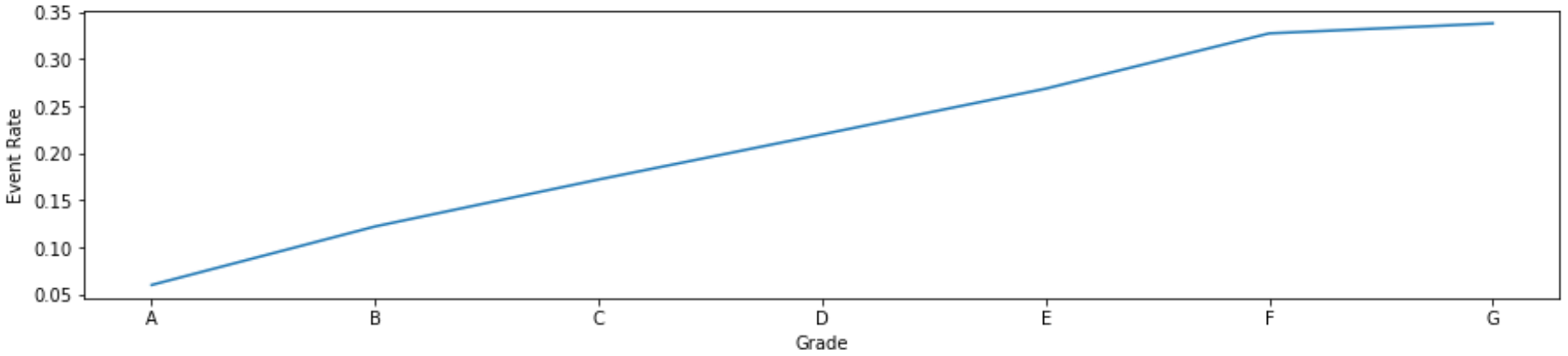
Cutoff	N	Events	% of Events	Non-Events	% of Non-Events	WoE	IV	Event_Percentage	Cutoff_Min
(5.419, 7.14]	4186	196	0.034832	3990	0.121221	1.24708	0.107734	0.046823	7.14
(7.14, 7.9]	3818	233	0.041407	3585	0.108917	0.967123	0.06529	0.061027	7.9
(7.9, 9.99]	4042	370	0.065754	3672	0.11156	0.528637	0.024215	0.091539	9.99
(9.99, 10.99]	4256	485	0.086192	3771	0.114568	0.284595	0.008076	0.113957	10.99
(10.99, 11.71]	2990	424	0.075351	2566	0.077958	0.034018	0.000089	0.141806	11.71
(11.71, 12.84]	3938	596	0.105918	3342	0.101534	-0.042268	0.000185	0.151346	12.84
(12.84, 13.8]	3862	612	0.108761	3250	0.098739	-0.096674	0.000969	0.158467	13.8
(13.8, 15.21]	3777	699	0.124222	3078	0.093514	-0.283967	0.00872	0.185068	15.21
(15.21, 16.82]	3920	850	0.151057	3070	0.093271	-0.482156	0.027862	0.216837	16.82
(16.82, 24.4]	3753	1162	0.206504	2591	0.078718	-0.964451	0.123244	0.309619	24.4



Most important variable is coming as Int_Rate.
But at the time of acquisition we give higher interest in two scenarios.
1. When customer is asking for more loan and we want to give the same amount but as per current process he/she is eligible for small ticket size.
2. When customer chances of default is little high as compare to other population.
Now as we have already given the loan to customer, risk assessment which we have taken into consideration at the time of acquisition is holding true.

Grade

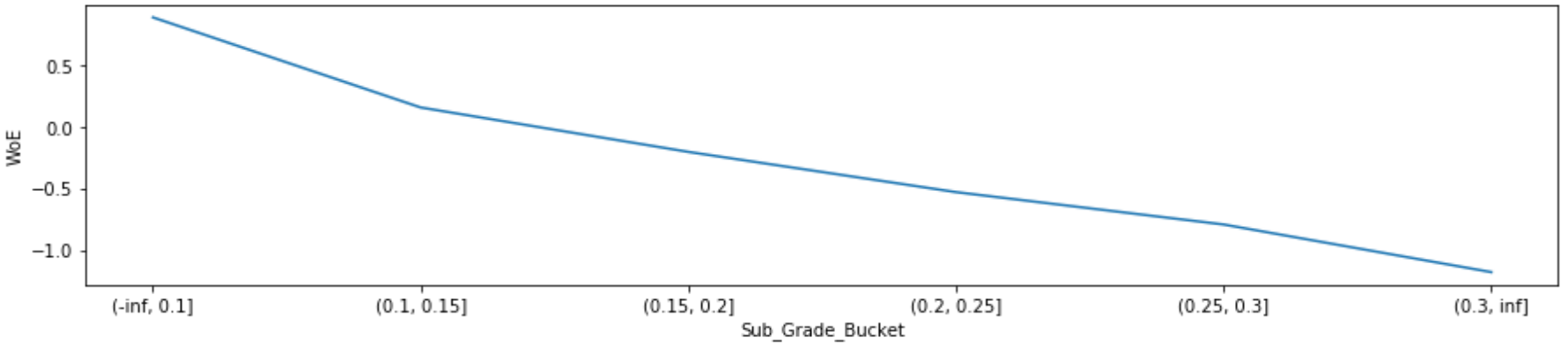
Cutoff	N	Events	% of Events	Non-Events	% of Non-Events	WoE	IV	Event_Percentage	Cutoff_Min
A	10027	602	0.106984	9425	0.286344	0.984512	0.176581	0.060038	A
B	11664	1425	0.253243	10239	0.311074	0.20568	0.011895	0.122171	B
C	7831	1347	0.239382	6484	0.196992	-0.194894	0.008261	0.172009	C
D	5083	1118	0.198685	3965	0.120462	-0.500388	0.039142	0.219949	D
E	2663	715	0.127066	1948	0.059183	-0.764076	0.051868	0.268494	E
F	975	319	0.056691	656	0.01993	-1.045382	0.038429	0.327179	F
G	299	101	0.017949	198	0.006015	-1.093206	0.013046	0.337793	G



Second Most important variable is coming as Grade.
We can not take any call on grade specifically as its just a type of a loan.
But at the time of collection we need to put more effort on grade 'G' customer as they have high chance of default as
compare to 'A' grade.

Sub Grade Bucket

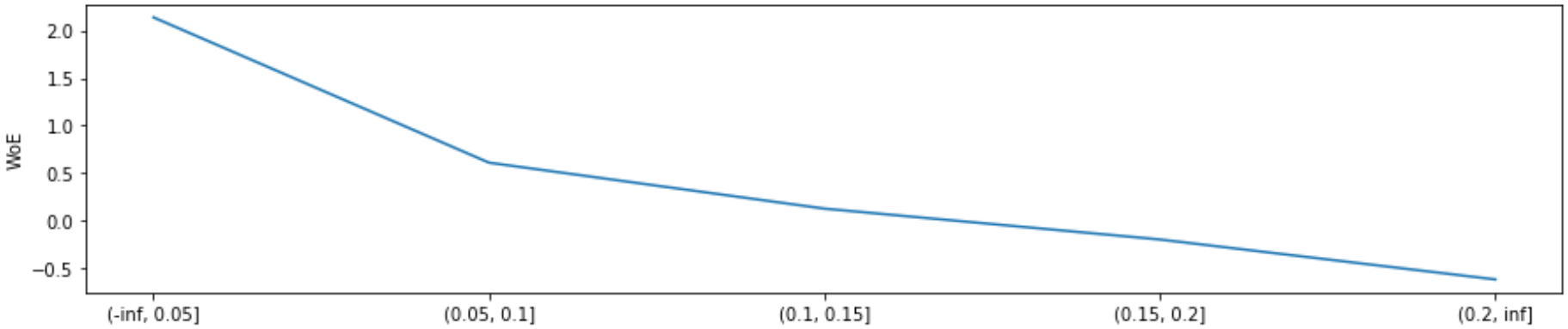
Cutoff	N	Events	% of Events	Non-Events	% of Non-Events	WoE	IV	Event_Percentage	Cutoff_Min
(-inf, 0.1]	11823	773	0.137373	11050	0.335713	0.893555	0.177227	0.065381	0.1
(0.1, 0.15]	9868	1254	0.222854	8614	0.261704	0.160698	0.006243	0.127077	0.15
(0.15, 0.2]	8761	1514	0.26906	7247	0.220173	-0.20052	0.009803	0.172811	0.2
(0.2, 0.25]	3890	874	0.155323	3016	0.09163	-0.527746	0.033613	0.224679	0.25
(0.25, 0.3]	3458	947	0.168296	2511	0.076287	-0.791215	0.072798	0.273858	0.3
(0.3, inf]	742	265	0.047094	477	0.014492	-1.178565	0.038424	0.357143	inf



Third Most important variable is coming as Sub Grade.
We can not take any call on sub grade specifically as its just a sub type of a Loan.
But at the time of collection we need to put more effort based on sub grade.
Sub Grade :- F2,G1,G5,F4,G2,G3,F5 need to focus on these category while collection. As chances of default is high.

Zip code

Cutoff	N	Events	% of Events	Non-Events	% of Non-Events	WoE	IV	Event_Percentage	Cutoff_Min
(-inf, 0.05]	1314	26	0.004621	1288	0.039131	2.136397	0.073728	0.019787	0.05
(0.05, 0.1]	6545	556	0.098809	5989	0.181954	0.610559	0.050764	0.08495	0.1
(0.1, 0.15]	15578	2037	0.362005	13541	0.411393	0.127892	0.006316	0.130761	0.15
(0.15, 0.2]	9135	1573	0.279545	7562	0.229743	-0.196201	0.009771	0.172195	0.2
(0.2, inf]	5970	1435	0.25502	4535	0.137779	-0.615692	0.072185	0.240369	inf

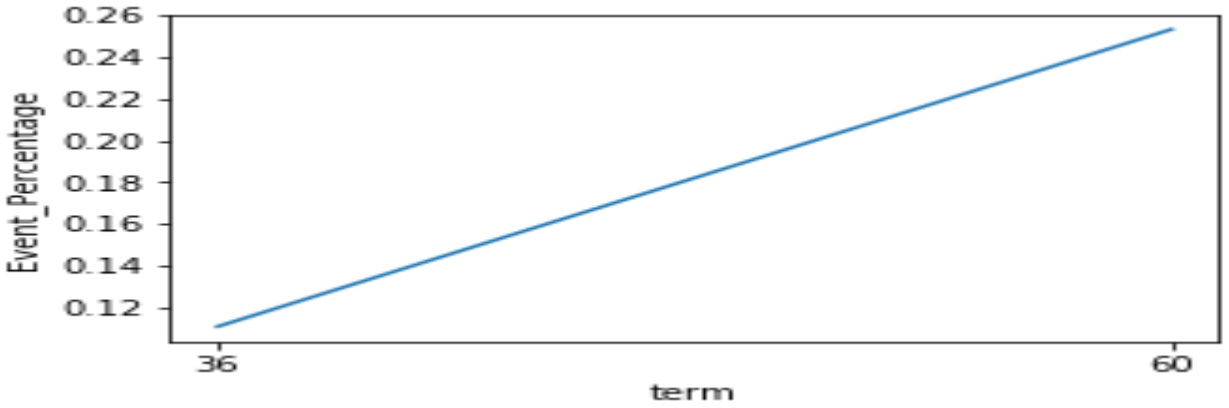


Forth Most important variable is coming as Zip-Code.
We can not take any call on zip code specifically as its just a zip code. We can't say if you are coming from this zip code we will no give loan to you.
But at the time of collection we need to put more effort based on zip code.

Zip code :- Need to focus on below list of zip code as they have high chance of default. '871xx', '346xx', '925xx', '751xx', '630xx', '339xx', '488xx','330xx', '245xx', '207xx', '905xx', '914xx', '333xx', '331xx','234xx', '917xx', '305xx', '571xx', '628xx', '171xx', '128xx','657xx', '341xx', '148xx', '960xx', '347xx', '065xx', '446xx','633xx', '360xx', '400xx', '754xx', '816xx', '119xx', '014xx','434xx', '344xx', '279xx', '285xx', '983xx', '547xx', '616xx','172xx', '161xx', '037xx', '974xx', '984xx', '907xx', '363xx','081xx', '971xx', '856xx', '844xx', '489xx', '013xx', '312xx','271xx', '302xx', '927xx', '444xx', '265xx', '238xx', '108xx','361xx', '321xx', '349xx', '641xx', '906xx', '890xx', '623xx','259xx', '779xx', '106xx', '675xx', '325xx', '497xx', '795xx','725xx', '154xx', '135xx', '593xx', '671xx', '487xx', '635xx','484xx', '546xx', '745xx', '449xx', '392xx', '056xx', '407xx', '367xx', '976xx', '937xx', '278xx', '354xx', '445xx', '439xx','072xx', '559xx', '283xx', '308xx', '153xx', '891xx', '082xx','147xx', '206xx', '224xx', '986xx', '614xx', '711xx', '626xx','447xx', '863xx', '826xx', '158xx', '615xx', '422xx', '244xx','619xx', '997xx', '026xx', '187xx', '177xx', '570xx', '534xx','075xx', '935xx', '766xx', '807xx', '758xx', '883xx', '035xx','859xx', '638xx', '425xx', '724xx', '264xx', '713xx', '316xx','376xx', '599xx', '668xx', '406xx', '409xx', '639xx', '912xx','924xx', '253xx', '611xx', '499xx', '573xx', '755xx', '438xx','215xx', '808xx', '744xx', '719xx', '203xx', '608xx', '607xx','897xx', '413xx', '798xx', '673xx', '496xx', '416xx','685xx','746xx', '561xx', '833xx', '663xx', '385xx', '669xx', '373xx','689xx', '094xx', '999xx'

Term

Cutoff	N	Events	% of Events	Non-Events	% of Non-Events	WoE	IV	Event_Percentage	Cutoff_Min
36	29063	3227	0.573485	25836	0.784931	0.313864	0.066365	0.111035	36
60	9479	2400	0.426515	7079	0.215069	-0.684688	0.144774	0.253191	60

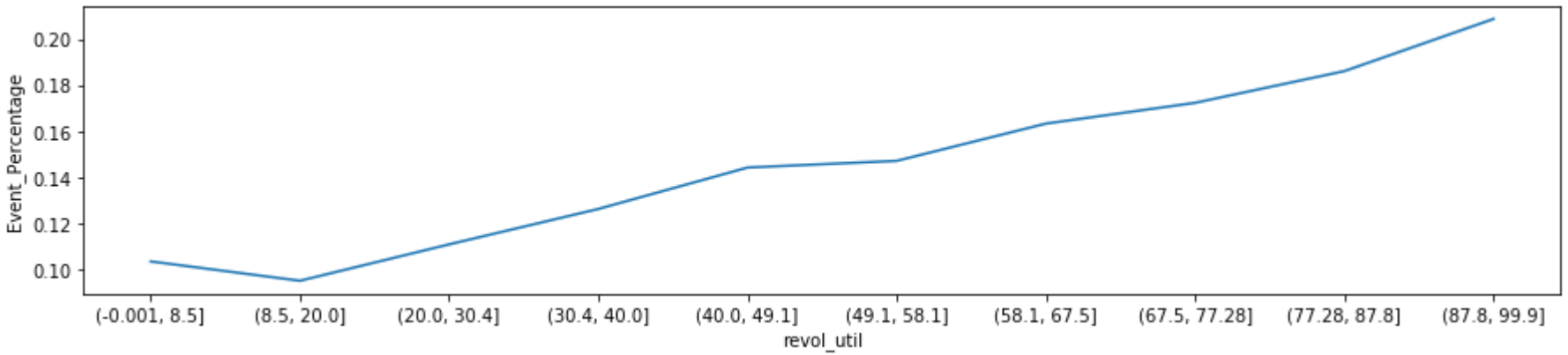


Fifth Most important variable is coming as Term.
We can not take any call on term as higher the tenure higher the profit but in return higher the chances of default.
But at the time of collection we need to put more effort based on term.
Term :- Customer taking loan for 60 Months have high chance of default. Need to focus at the time of collection.

SUGGESTIONS (AT THE TIME OF ACQUISITION)

Revol_Util

Cutoff	N	Events	% of Events	Non-Events	% of Non-Events	WoE	IV	Event_Percentage	Cutoff_Min
(-0.001, 8.5]	3875	402	0.071441	3473	0.105514	0.38997	0.013287	0.103742	8.5
(8.5, 20.0]	3860	368	0.065399	3492	0.106091	0.483795	0.019687	0.095337	20
(20.0, 30.4]	3851	428	0.076062	3423	0.103995	0.312797	0.008737	0.11114	30.4
(30.4, 40.0]	3849	487	0.086547	3362	0.102142	0.165675	0.002584	0.126526	40
(40.0, 49.1]	3875	560	0.09952	3315	0.100714	0.011924	0.000014	0.144516	49.1
(49.1, 58.1]	3819	563	0.100053	3256	0.098921	-0.011377	0.000013	0.147421	58.1
(58.1, 67.5]	3875	634	0.112671	3241	0.098466	-0.134764	0.001914	0.163613	67.5
(67.5, 77.28]	3829	661	0.117469	3168	0.096248	-0.19925	0.004228	0.17263	77.28
(77.28, 87.8]	3868	721	0.128132	3147	0.09561	-0.292786	0.009522	0.186401	87.8
(87.8, 99.9]	3841	803	0.142705	3038	0.092298	-0.435752	0.021965	0.20906	99.9

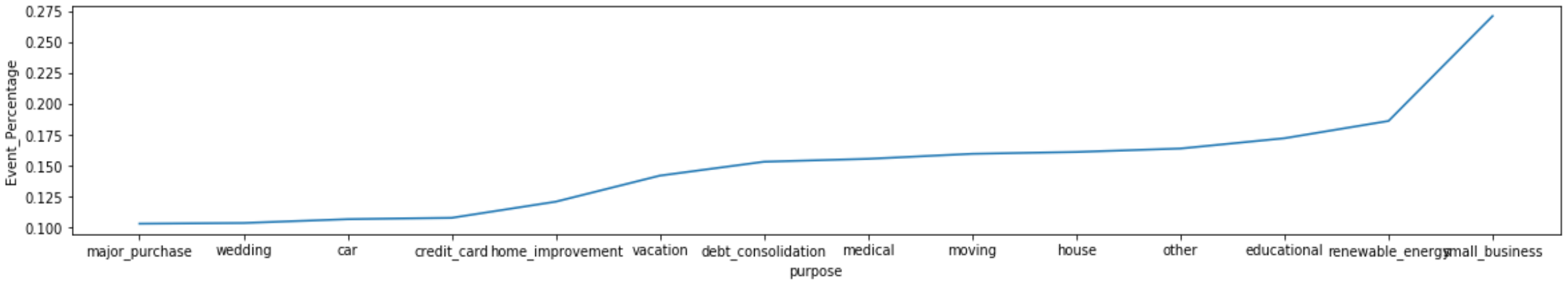


sixth Most important variable is coming as revol_util.
We can take a call on this. We can use this variable at the time of acquisition.
customer with higher revol_util tends to do more default so we can give higher interest rate or not give the loan at all.
revol_util :- Customer with higher revol_util tends to do more default.

Purpose

Cutoff	N	Events	% of Events	Non-Events	% of Non-Events	WoE	IV	Event_Percentage	Cutoff_Min
major_purchase	2150	222	0.039453	1928	0.058575	0.395209	0.007557	0.103256	major_purchase
wedding	925	96	0.017061	829	0.025186	0.38952	0.003165	0.103784	wedding
car	1497	160	0.028434	1337	0.04062	0.356658	0.004346	0.10688	car
credit_card	5020	542	0.096321	4478	0.136047	0.345314	0.013718	0.107968	credit_card
home_improvement	2867	347	0.061667	2520	0.076561	0.216337	0.003222	0.121032	home_improvement
vacation	373	53	0.009419	320	0.009722	0.031677	0.00001	0.142091	vacation
debt_consolidation	18048	2767	0.491736	15281	0.464256	-0.057506	0.00158	0.153313	debt_consolidation
medical	681	106	0.018838	575	0.017469	-0.075421	0.000103	0.155653	medical
moving	576	92	0.01635	484	0.014705	-0.106056	0.000174	0.159722	moving
house	366	59	0.010485	307	0.009327	-0.117042	0.000136	0.161202	house
other	3860	633	0.112493	3227	0.09804	-0.137514	0.001987	0.16399	other
educational	325	56	0.009952	269	0.008173	-0.196992	0.000351	0.172308	educational
renewable_energy	102	19	0.003377	83	0.002522	-0.29195	0.00025	0.186275	renewable_energy
small_business	1752	475	0.084414	1277	0.038797	-0.777398	0.035463	0.271119	small_business

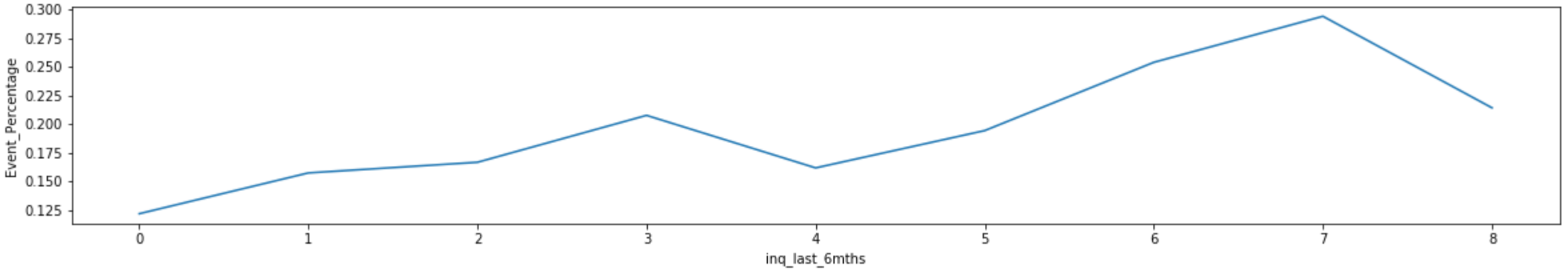
Seventh Most important variable is coming as purpose.
We can take a call on this. We can use this variable at the time of acquisition.
Customer giving purpose as small business tends to do more default.
purpose :- Purpose as small business customer tends to default more.



Inquiry Last 6 Months

Cutoff	N	Events	% of Events	Non-Events	% of Non-Events	WoE	IV	Event_Percentage	Cutoff_Min
0	18694	2280	0.405189	16414	0.498678	0.207607	0.019409	0.121964	0
1	10645	1677	0.298027	8968	0.272459	-0.089696	0.002293	0.157539	1
2	5653	943	0.167585	4710	0.143096	-0.157975	0.003869	0.166814	2
3	2980	619	0.110005	2361	0.07173	-0.427617	0.016367	0.207718	3
4	315	51	0.009063	264	0.008021	-0.122229	0.000127	0.161905	4
5	144	28	0.004976	116	0.003524	-0.344966	0.000501	0.194444	5
6	63	16	0.002843	47	0.001428	-0.688793	0.000975	0.253968	6
7	34	10	0.001777	24	0.000729	-0.890883	0.000934	0.294118	7
8	14	3	0.000533	11	0.000334	-0.467069	0.000093	0.214286	8

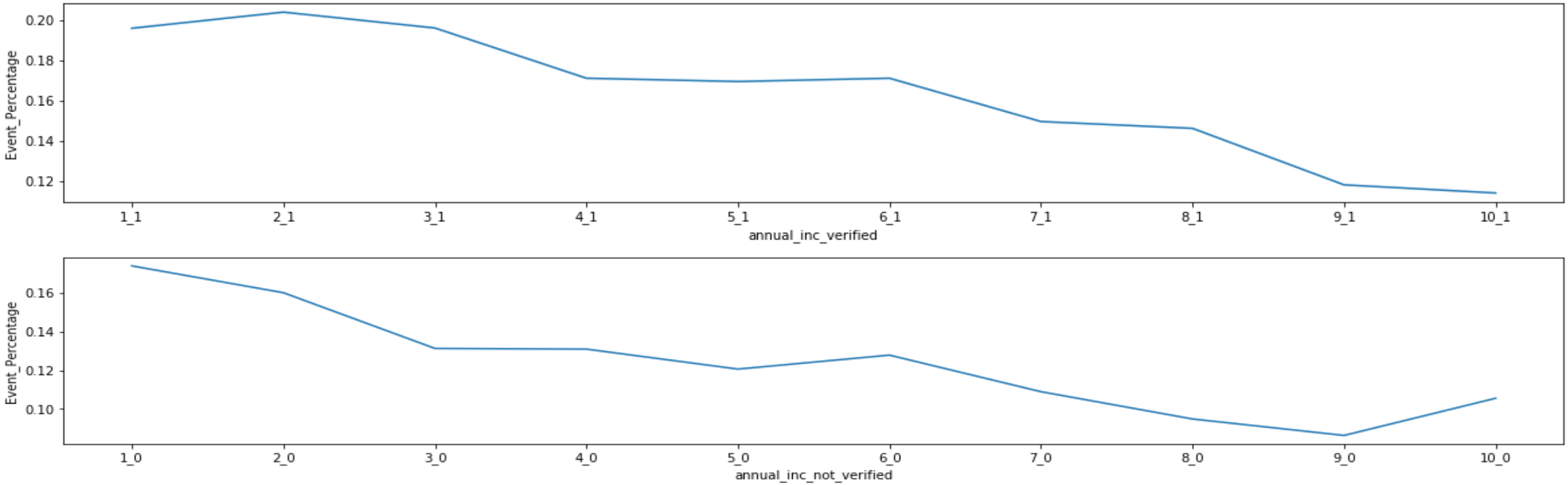
Eighth Most important variable is coming as inq_last_6mths.
We can take a call on this. We can use this variable at the time of acquisition.
Higher the number of inquire higher the chance of default.
This is logical as customer is not getting loan in the market that's why number of inquiry is high.
inq_last_6mths :- Higher the number of inquire higher the chance of default.



Annual Income With Verification Status

Cutoff	N	Events	% of Events	Non-Events	% of Non-Events	WoE	IV	Event_Percentage	Cutoff_Min
10_0	958	101	0.017949	857	0.026037	0.371965	0.003008	0.105428	0
1_0	2417	421	0.074818	1996	0.060641	-0.210084	0.002978	0.174183	0
8_0	1542	146	0.025946	1396	0.042412	0.491408	0.008091	0.094682	0
2_0	1610	258	0.04585	1352	0.041075	-0.109971	0.000525	0.160248	0
3_0	2018	265	0.047094	1753	0.053258	0.123002	0.000758	0.131318	0
7_0	1571	171	0.030389	1400	0.042534	0.336212	0.004083	0.108848	0
4_0	1909	250	0.044429	1659	0.050403	0.126157	0.000754	0.130959	0
9_0	1219	105	0.01866	1114	0.033845	0.5954	0.009041	0.086136	0
5_0	1700	205	0.036431	1495	0.04542	0.220519	0.001982	0.120588	0
6_0	1721	220	0.039097	1501	0.045602	0.153907	0.001001	0.127833	0
8_1	2491	364	0.064688	2127	0.064621	-0.001038	6.97E-08	0.146126	1
7_1	2321	347	0.061667	1974	0.059973	-0.02786	4.72E-05	0.149505	1
6_1	2151	368	0.065399	1783	0.05417	-0.188382	0.002115	0.171083	1
4_1	1999	342	0.060778	1657	0.050342	-0.188399	0.001966	0.171086	1
3_1	1846	362	0.064333	1484	0.045086	-0.3555	0.006842	0.1961	1
2_1	1549	316	0.056158	1233	0.03746	-0.404889	0.00757	0.204003	1
1_1	2133	418	0.074285	1715	0.052104	-0.354665	0.007867	0.195968	1
10_1	2879	328	0.05829	2551	0.077503	0.284875	0.005473	0.113928	1
5_1	2101	356	0.063266	1745	0.053015	-0.176773	0.001812	0.169443	1
9_1	2407	284	0.050471	2123	0.064499	0.245259	0.003441	0.117989	1

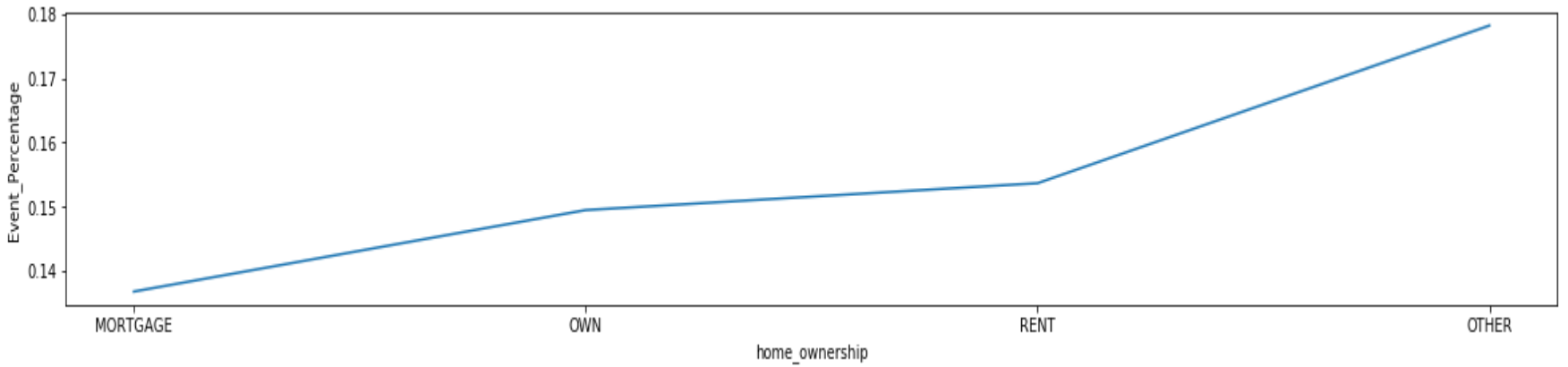
Annual Income With Verification Status (Continue)



Ninth Most important variable is coming as annual_inc with Verification Status.
Higher the annual_inc lower the chance of default.
IV of annual_income with verification status is higher than normal annual income
annual income with reference to verification status :- Higher the income in both cases higher the chance of default.

Home Ownership

Cutoff	N	Events	% of Events	Non-Events	% of Non-Events	WoE	IV	Event_Percentage	Cutoff_Min
MORTGAGE	17006	2327	0.413542	14679	0.445967	0.075486	0.002448	0.136834	MORTGAGE
OWN	2963	443	0.078728	2520	0.076561	-0.027908	0.00006	0.149511	OWN
RENT	18472	2839	0.504532	15633	0.474951	-0.06042	0.001787	0.153692	RENT
OTHER	101	18	0.003199	83	0.002522	-0.237883	0.000161	0.178218	OTHER



Tenth Most important variable is coming as home_ownership.
We can take a call on this. We can use this variable at the time of acquisition.
Customer not owning property tends to do more default.
home_ownership :- Home ownership as Other tends to do more default.

TWO WAYS WE CAN REDUCE THE RISK.

1.By focusing on customer at time of collection.

- Int_Rate :- Need to focus on collection and put more effort on customer having higher interest rate.
- Grade :- Grade G have higher chances of default need to focus on these type of customer more.
- Sub Grade :- F2,G1,G5,F4,G2,G3,F5 need to focus on these category while collection. As chances of default is high.
- Term :- Customer taking loan for 60 Months have high chance of default. Need to focus at the time of collection.
- Zip code :- Need to focus on below list of zip code as they have high chance of default.

'871xx', '346xx', '925xx', '751xx', '630xx', '339xx', '488xx','330xx', '245xx', '207xx', '905xx', '914xx', '333xx', '331xx','234xx', '917xx', '305xx', '571xx', '628xx', '171xx', '128xx','657xx', '341xx', '148xx', '960xx', '347xx', '065xx', '446xx','633xx', '360xx', '400xx', '754xx', '816xx', '119xx', '014xx','434xx', '344xx', '279xx', '285xx', '983xx', '547xx', '616xx','172xx', '161xx', '037xx', '974xx', '984xx', '907xx', '363xx','081xx', '971xx', '856xx', '844xx', '489xx', '013xx', '312xx','271xx', '302xx', '927xx', '444xx', '265xx', '238xx', '108xx','361xx', '321xx', '349xx', '641xx', '906xx', '890xx', '623xx','259xx', '779xx', '106xx', '675xx', '325xx', '497xx', '795xx','725xx', '154xx', '135xx', '593xx', '671xx', '487xx', '635xx','484xx', '546xx', '745xx', '449xx', '392xx', '056xx', '407xx', '367xx', '976xx', '937xx', '278xx', '354xx', '445xx', '439xx','072xx', '559xx', '283xx', '308xx', '153xx', '891xx', '082xx','147xx', '206xx', '224xx', '986xx', '614xx', '711xx', '626xx','447xx', '863xx', '826xx', '158xx', '615xx', '422xx', '244xx','619xx', '997xx', '026xx', '187xx', '177xx', '570xx', '534xx','075xx', '935xx', '766xx', '807xx', '758xx', '883xx', '035xx','859xx', '638xx', '425xx', '724xx', '264xx', '713xx', '316xx','376xx', '599xx', '668xx', '406xx', '409xx', '639xx', '912xx','924xx', '253xx', '611xx', '499xx', '573xx', '755xx', '438xx','215xx', '808xx', '744xx', '719xx', '203xx', '608xx', '607xx','897xx', '413xx', '798xx', '673xx', '496xx', '416xx','685xx','746xx', '561xx', '833xx', '663xx', '385xx', '669xx', '373xx','689xx', '094xx', '999xx‘

2.By Changing Acquisition Score model and including below variables.

- revol_util :- Customer with higher revol_util tends to do more default.
- purpose :- Purpose as small business customer tends to default more.
- inq_last_6mths :- Higher the number of inquire higher the chance of default.
- annual income with reference to verification status :- Higher the income in both cases higher the chance of default.
- homeownership :- Home ownership as Other tends to do more default.

*** Variable used in rejecting a customer are not used like amount requested, application date, loan title, risk score, DTI, zip code, State, Emp Length, Policy Code