Lending Club Case Study

Problem statement

- We would be analysing the data collected from a consumer finance company regarding the details of people for which loans were approved in the time period of 2007 to 2011.
- There are situations where people fully pay the loan, are in the process of paying (current), or did not complete the payment on time (default).
- The defaulted applicants cause a major loss in this sector, and so the aim of this study is to analyse the dataset and identify the factors that are responsible that could cause people to default.
- There could be various driving factors to this, likely related to the annual income, total amount to be paid, work experience etc.
- We will be doing univariant and bivariant analysis on the data to come up with relevant information that could be useful for the company for future risk assessment.

I. <u>Data Cleaning</u>

- The contents of the dataset needs to be cleaned up to add/update relevant column names, remove columns which won't be useful in the analysis and having the dataset in the format that would be useful to do proper analysis on.
- Based on the type of data seen from the file, the below fields are being removed, as it doesn't provide any useful info related to our problem statement:
 - > url: irrelevant for analysis
 - **desc**: irrelevant for analysis
 - > pymnt_plan: same for all applicants
 - > initial list status: same for all applicants
 - > collections_12_mths_ex_med total_il_high_credit_limit (rest all columns): these are either same for all or are NA
- We also remove the columns that have more than 20,000 records missing data.
- Some columns are updated with rounding the decimal points to 2 characters.
- It is checked if there are any duplicate records. If so, they will be removed.
- After these steps are completed, we are left with a cleaner, smaller dataset.

II. <u>Univariant Analysis</u>

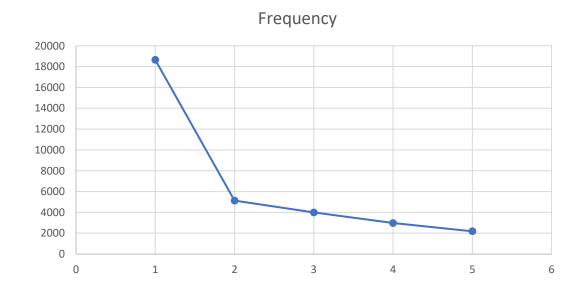
- The dataset contains various types of columns that could be falling into either of the 2 variables:
 - > Ordered: The set of columns that have a particular ordering. Some examples are:
 - o emp_length
 - o grade
 - o sub_grade
 - > Unordered: Ones that do not have a particular order to be classified in. Some examples are:
 - o home ownership
 - verification_status
 - o purpose
- Apart from these, we also have quantitative variables here. For example:
 - loan_amnt
 - o int rate

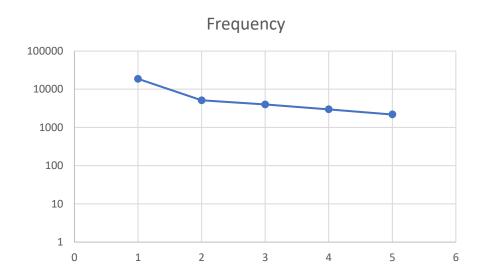
Let's try to see the frequency plot for some of the unordered sets

Row Labels	Count of home_ownership
MORTGAGE	17659
NONE	3
OTHER	98
OWN	3058
RENT	18899
Grand Total	39717

The first plot shows the frequency vs rank scatter plot based on ranking the home ownership of more counts.

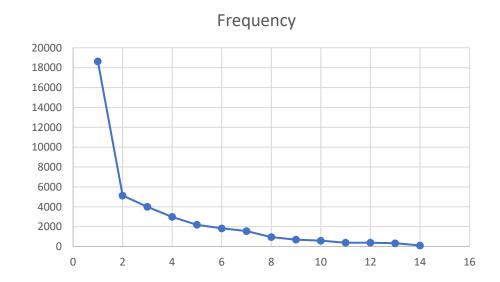
The second plot shows the same distribution but with Y axis in log scale. This shows us that the curve is turning to be a straight line.

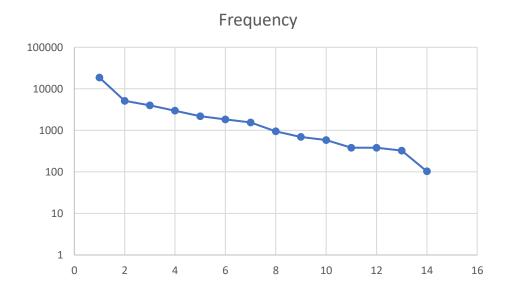




Row Labels	Count of purpose
debt_consolidation	18641
credit_card	5130
other	3993
home_improvement	2976
major_purchase	2187
small_business	1828
car	1549
wedding	947
medical	693
moving	583
vacation	381
house	381
educational	325
renewable_energy	103
Grand Total	39717

Similar to the previous example, we can see that the curves are turning more into a straight line as we move into a log scale





Let's try to see the frequency plot for some of the ordered sets

Row Labels	Count of emp_length	
10+ years		8879
< 1 year		4583
2 years		4388
3 years		4095
4 years		3436
5 years		3282
1 year		3240
6 years		2229
7 years		1773
8 years		1479
9 years		1258
n/a		1075
Grand Total		39717



This analysis shows us that the people in the employment length of more than 10 years are more likely to opt for the loan than the other categories

The different descriptive statistics for Quantitative variables are done

- Mean
- Median
- Interquartile difference

loan	amnt
-cui	aiiiii

5000	
2500	
2400	
10000	
3000	
5000	
7000	
3000	
5600	
5375	
6500	
12000	

9000

3000 10000

1000

Mean	Median	SD	Quantities	Values
11219.44	10000	7456.671	25th percentile	5500
			50th percentile	10000
			75th percentile	15000
			100th percentile	35000

Here, we can see that the quartiles are a good way to understand the spread. The loan amount would be a big factor to consider in this analysis. In order to get the proper set of details, it would be good to analyse the data between the 25th percentile and the 75th percentile.

III. Segmented Univariant Analysis

Grade	Charged Off	Current F	G Fully Paid T	irand otal
Α	602	40	9443	10085
В	1425	345	10250	12020
С	1347	264	6487	8098
D	1118	222	3967	5307
Е	715	179	1948	2842
F	319	73	657	1049
G	101	17	198	316
Grand Total	5627	1140	32950	39717

Experience	Charged Off		Current	Fully Paid	Grand Total
< 1 year		639	75	3869	4583
1 year		456	71	. 2713	3240
10+ years		1331	391	. 7157	8879
2 years		567	97	3724	4388
3 years		555	83	3457	4095
4 years		462	94	2880	3436
5 years		458	88	2736	3282
6 years		307	61	1861	. 2229
7 years		263	62	1448	1773
8 years		203	44	1232	1479
9 years		158	32	1068	1258
n/a		228	42	805	1075
Grand Total		5627	1140	32950	39717

Home ownership	Charged Off	Cu	ırrent Ful	lly Paid	Grand Total
MORTGAGE		2327	638	14694	17659
NONE				3	3
OTHER		18		80	98
OWN		443	83	2532	3058
RENT		2839	419	15641	18899
Grand Total		5627	1140	32950	39717

The analysis has been done for various factors against the loan status. The major ones that could be pointed out here are Home ownership and Work Experience. It seems like people living on Rent, as well as people having more than 10 years of exp have more chances of defaulting.

IV. <u>Bivariant Analysis</u>

Row Labels	Average of annual_inc
car	61842.04167
Charged Off	54560.0245
Current	58038.82
Fully Paid	62854.20285
credit_card	70439.14779
Charged Off	64052.04421
Current	75787.81456
Fully Paid	71088.17733
debt_consolidation	67322.05922
Charged Off	61665.68666
Current	74824.57218
Fully Paid	68058.2386
educational	53471.37409
Charged Off	51711.80357
Fully Paid	53837.67874
home_improvement	89736.78495
Charged Off	77190.18882
Current	9655.15406
Fully Paid	91186.55297
house	76772.28388
Charged Off	71540.46847
Current	77157.14286
Fully Paid	77137.14280
	66391.5229
major_purchase	56707.54523
Charged Off	
Current	59994.16757
Fully Paid	67629.35755
medical	68252.86377
Charged Off	57261.35849
Current	111103.08
Fully Paid	69384.85849
moving	61801.5783
Charged Off	55533.76087
Current	47465.28571
Fully Paid	63200.32469
other	63147.25236
Charged Off	58676.59368
Current	72583.74844
Fully Paid	63649.12595
renewable_energy	77490.00612
Charged Off	59240.82263
Current	109000
Fully Paid	81287.89157
small_business	75062.51516
Charged Off	67556.26162
Current	71143.43243
Fully Paid	78076.96594
vacation	59218.93346
Charged Off	52452.5283
Current	65000
Fully Paid	60224.9368
wedding	68663.28147
Charged Off	66634.81542
Current	79228.11905
	68630.59611
Fully Paid	68630,59611

Different set of analysis have been conducted and analysed.

This is an example of the details related to the purpose of the loan and how that affects the loan status, based on average annual income of the person as well.

	Row Labels	Charged Off	Current	Fully Paid	Grand Total
	<1 year	6	39	75	3869 4583
□ 1988 日	MORTGAGE	1	77	39	1085 1301
OW 32 4 30 30 20 20 20 20 20 20 20 30 30 20 30 </td <td>NONE</td> <td></td> <td></td> <td></td> <td>2 2</td>	NONE				2 2
sate type 68 70 20 38 70 20 38 70 70 38 70	OTHER		2		19 21
Speed 456 72 222 1246 294 ORMORDIGIG 14 1 1 15 5 15 15 5 125 254 254 254 254 254 254 254 252	OWN		52	4	288 344
MOTRIGATION 11 30 50 10 13 30 10 10 30 20	RENT	4	08	32	2475 2915
MOTHORISM 137 30 106 579 500 51 15 55 520 500 200 </td <td>1 year</td> <td>4</td> <td>56</td> <td>71</td> <td>2713 3240</td>	1 year	4	56	71	2713 3240
OMM 18 11 16 28 13 136 283		1	37	30	806 973
OW 28 4 156 231 235 233 237 238 238 238 237 238 238 238 238 235	OTHER		4		11 15
1949	OWN		28	4	
1949	RENT	2	87	37	1710 2034
Montack7326435933Onthe1924132155Inter19717174184Street19717174184Contract121125184Contract121125120Contract121125120Contract133213120Contract13321120Contract13110120Contract13710120Contract13710120Contract13710120Contract13710120Contract13710120Contract12110120Contract12110120Contract12110120Contract121112Contract12111Contract12111Contract12111Contract12111Contract12111Contract12111Contract12111Contract12111Contract12111Contract12111Contract12 <td< td=""><td></td><td></td><td></td><td></td><td></td></td<>					
ONIM 5 14 13 55 REAT 474 103 1575 2575 2574 3585 156 355 126 3585 126 3585 126 3585 126 3585 126 3585 126					
OWN 196 24 102 153 255 Sters 197 197 1978 1981 1981 CONTINA 12 40 125 180 100 OTHAR 12 4 26 18 10 OWN 13 4 25 18 20 20 BERT 136 33 2255 18 387 489 389 MORTINAGE 13 3 17 10					
BRIT 454 10 1055 255 257 30 358 388 <td>OWN</td> <td></td> <td>99</td> <td>24</td> <td></td>	OWN		99	24	
Step 57 75 374 488 488 186					
MORTRAGE 174 40 125 140 ONN 43 4 226 28 RINT 48 53 225 265 Seen 58 8 347 488 MORTRAGE 20 35 134 197 100 ONA 3 7 208 235 RINT 31 41 1988 276 208 33 RINT 42 41 1988 226 34 1988 226 34 1988 226 34 1988 226 34 1988 226 34 1988 226 34 1988 226 34 1988 226 34 1988 226 34 1988 226 34 1988 226 34 1988 226 34 198 226 124 198 226 124 198 226 128 128 128 128 128 128 <td></td> <td></td> <td></td> <td></td> <td></td>					
の					
OWN 43 4 255 225 2825 SHOT 355 31 397 485					
1987				4	
Symmeth 555 13 4957 4958 MORITGAGE 22 2 5 134 177 DINES 3 7 20 150 OWN 31 7 208 252 INSTANT 312 41 1508 252 INSTANT 42 42 1508 252 INSTANT 16 43 1518 1502 ONTEAG 17 7 7 7 OWN 28 42 160 27 INTERCAC 158 42 160 170 INTERCAC 158 42 160 170 INTERCAC 159 42 160 170 INTERCAC 15 4 15 15 INTERCAC 15 4 15 15 INTERCAC 15 12 13 15 INTERCAC 15 12 13 15					
DOTTICATE 130 131 1371					
□ DTHEN					
OWN 38 7 208 233 RENT 312 41 1068 228 MONTGAGE 49 48 188 1402 OTHER T 7 7 7 OWN 28 9 204 244 244 RENT 28 49 1048 276 248 SYREN 458 48 276 128 148 MONTGAGE 19 49 128 148 128 148 128 148 128 148 128 148		•		33	
IRENT				7	
4yers 462 34 280 348 MORTGAGE 16 43 133 160 OTHER 7 7 7 7 OWN 28 9 204 218 ERIT 258 42 126 1276 Syens 458 8 276 328 MORTGAGE 17 4 128 1 4 1					
MORTGAGE 176 43 1183 1402 OTHER 7 7 7 7 7 7 7 7 7 7 7 7 7 128 28 9 20 20 218 128					
OTHER T T OWN 28 9 204 218 SYERT 458 88 2796 3282 NORTGAGE 197 49 1238 1238 NORE 197 49 1238 138 NOR 197 49 1238 138 NOR 36 7 190 233 SWIT 224 32 1303 159 SWIT 24 32 1303 159 SWIT 24 32 1303 159 SWIT 36 7 190 233 SWIT 36 7 190 233 SWIT 37 19 25 1002 SWIT 38 39 31 31 36 SWIT 49 22 30 31 31 31 31 31 31 31 31 31 32 33 33					
OWN 28 9 204 241 RENT 258 42 1486 1286 Sysers 458 88 2736 2322 MORTGAGE 197 49 1238 1484 OTHER 1 1 4 5 OWN 36 7 190 233 SRENT 224 32 1303 1559 Syers 307 61 1861 2229 OTHER 1 4 3 159 Syers 307 61 1861 2229 OTHER 1 4 4 4 OWN 27 3 131 161 4 RENT 149 22 301 972 3 131 161 192 192 193 184 1773 184 1773 184 1773 184 1773 184 1773 184 1173 184 1174		•	70	43	
RENT 258 42 1486 2786 Syears 458 88 2796 3282 MORITGAGE 197 49 1138 1484 NOME 1 2 1 2 1 1 2 1 1 2 1 1 4 4 4 4 4 4 4 4 4 1 4 1 3 4			20	0	
Syenry 468 88 278 2828 MORTGAGE 197 49 1238 1484 MORTGAGE 1 1 1 1 1 OTHER 1 1 4 5 5 OWN 36 7 109 233 153 SRET 24 32 1303 1593 1593 Syears 307 61 1861 2229 MORTGAGE 131 36 925 1092 OTHER 149 23 331 161 ENT 149 22 33 131 161 ENT 149 23 43 131 161 ENT 149 23 44 133 884 OTHER 1 2 4 19 173 MORTGAGE 11 3 4 19 14 OWN 13 4 13 14 14 <td></td> <td></td> <td></td> <td></td> <td></td>					
MORTGAGE 197 49 1288 1484 NONE 1					
NONE 1 4 5 OTHER 1 4 5 OWN 36 7 190 233 RENT 224 32 1303 1559 5 years 307 61 1861 229 OTHER 4 6 24 2 1002 OWN 27 3 331 161 4 WAR 27 3 331 161 922 Years 263 62 148 173 84 MORTGAGE 117 3 73 33 88 173 3 88 173 88 173 88 173 88 173 88 174 140					
OTHER 1 4 5 OWN 36 7 190 233 SPENT 224 32 1033 1559 SPENT 307 61 1861 2229 MORTGAGE 131 36 95 1092 OTHER 4 4 4 4 CWN 27 3 131 161 92 RENT 199 22 801 922 181 92 92 92 92 92 92 92 92 92		•	97	49	
OWN 36 7 190 233 RENT 224 32 1303 1559 Syears 307 61 1861 222 MORTGAGE 131 36 925 1092 OTHER 4 4 4 4 OWN 27 3 131 166 972 RENT 149 22 801 972					
RENT 224 32 1303 1559 6 years 307 61 1861 2229 MORTEAGE 131 36 925 1092 OTHER 4 4 4 OWN 27 3 131 161 RENT 149 22 801 972 Years 263 62 148 1773 MORTEAGE 117 34 733 884 OTHER 1 2 1 3 733 884 OTHER 1 3 733 884 177 3 184 1733 884 OTHER 1 2 4 14 140				7	
Syears 307 61 1861 2229 MORTGAGE 131 36 925 1092 OTHER 4 4 4 4 OWN 27 3 131 161 972 YEARS 263 62 148 1773 884 OTHER 17 3 73 884 1773 884 1773 884 1773 884 1773 884 1773 884 1773 884 1773 1884 1773 1884 1773 1884 1773 1884 1773 1884 1773 1884 1773 1884 1884 1884 1894 <td></td> <td></td> <td></td> <td></td> <td></td>					
MORTGAGE 131 36 925 1092 OTHER 4 4 4 OWN 27 3 131 161 RENT 149 22 801 972 7 ears 263 62 1448 1773 MORTGAGE 117 34 733 884 OWN 1 2 4 114 140 OWN 22 4 114 140 140 RENT 123 24 599 746 8 years 23 44 123 149 144 OTHER 102 23 556 781 14 140 14					
OTHER CTHER 4 4 OWN 27 3 131 161 RENT 149 22 801 922 Years 263 62 1448 1773 MORTEAGE 17 34 733 884 OTHER 1 2 4 144 140 RENT 123 24 599 746 8 Years 29 44 123 1479 OTHER 123 23 565 781 OTHER 123 24 159 1479 OWN 13 4 103 120 Years 18 17 469 574 Years 18 17 469 574 Years 18 17 469 128 MORTEAGE 19 23 600 702 OWN 14 1 77 92 RENT 15					
OWN 27 3 131 161 RENT 19 22 801 972 Years 263 62 1448 1773 MORTGAGE 117 34 733 884 OTHER 1 2 2 3 3 34 144 140 3 3 3 3 3 3 884 4 143 2 3 3 884 4 143 144 140 3 3 3 884 3 149 3 3 3 884 3 3 4 144 140 3 7 4		•	51	30	
RENT 149 22 801 972 Years 263 62 1448 1773 MORTGAGE 117 34 733 884 OTHER 1 2 3 12 3 3 14 11 14 <td></td> <td></td> <td>27</td> <td>2</td> <td></td>			27	2	
Tyears 263 62 1448 1773 MORTGAGE 117 34 733 884 OTHER 1 2 3 114 140 OWN 22 4 113 599 746 RENT 23 44 599 746 Syears 23 44 132 1479 OWN 13 4 103 120 Syears 13 4 103 120 Syears 13 4 103 120 Syears 18 17 469 574 Syears 18 17 469 574 Syears 18 17 469 72 MORTGAGE 19 23 600 702 OWN 14 1 77 92 WORTGAGE 28 3 39 46 MORTGAGE 28 21 34 447					
MORTGAGE 117 34 733 884 OTHER 1 2 2 3 34 114 140 140 141 140 140 141 140 140 141 140					
OTHER 1 2 3 OWN 22 4 114 140 RENT 123 24 59 746 8 years 203 44 1232 1479 MORTGAGE 102 23 656 781 OTHER 12 2 3 656 781 OWN 13 4 103 120 544 9 years 158 32 1068 1258 MORTGAGE 79 23 600 702 OWN 14 1 77 92 RENT 65 8 39 391 464 MORTGAGE 82 21 351 454 MORTGAGE 82 21 353 454 MORTGAGE 82 21 354 454 MORTGAGE 82 21 344 447 MORTGAGE 82 21 344 447					
OWN 22 4 114 140 RENT 123 24 599 746 8 years 23 44 1232 1379 MORTGAGE 102 23 556 781 OTHER 1 4 103 120 Years 13 4 103 120 9 years 158 32 1068 125 MORTGAGE 79 23 600 702 OWN 14 1 77 92 OWN 65 8 391 464 N/G 28 42 85 1075 MORTGAGE 82 21 34 447 OWN 43 12 34 447 OWN 43 12 34 447 0WN 43 12 34 447 0WN 3 12 34 447 10 3 9		•		34	
RENT 123 24 599 746 8 Years 203 44 1232 1479 OMDRTGAGE 102 23 656 781 OTHER 4 103 120 OWN 13 4 103 120 RENT 88 17 469 574 9years 158 32 108 1258 MORTGAGE 79 23 600 702 OWN 14 1 77 92 RENT 58 391 464 10/2 8 391 454 MORTGAGE 8 391 454 MORTGAGE 8 391 454 MORTGAGE 8 22 42 85 1075 MORTGAGE 8 21 34 44 47 OWN 3 12 34 44 47 QUN 3 12 34					
8 years 203 44 123 1279 MORTGAGE 102 23 656 781 OTHER 1 4 103 120 OWN 13 4 103 120 FERT 88 17 469 574 9 years 18 32 1068 1258 MORTGAGE 79 23 600 702 RENT 65 8 391 464 N/2 28 42 805 1075 MORTGAGE 8 21 343 447 OWN 8 21 345 447 OWN 8 21 345 427 GWY 9 29 217 411 RENT 103 9 29 411					
MORTGAGE 102 23 656 781 OTHER 4 4 4 4 OWN 13 4 103 120 574 574 499 574 499 574 499 574 499 574 499 574 498 574 498 574 498 572 588 400 702 702 702 702 702 702 702 702 702 702 702 703 702					
OTHER 4 4 4 4 103 120 OWN 13 4 103 120 574 459 574 459 574 459 128 42 168 1258 42 460 702 702 702 702 702 703 702 703 703 702 703 703 703 703 704 703 703 704 <t< td=""><td></td><td></td><td></td><td></td><td></td></t<>					
OWN 13 4 103 120 RENT 88 17 469 574 9 years 158 32 1068 1258 MORTGAGE 79 23 600 702 RENT 65 8 391 464 v/a 228 42 805 1075 MORTGAGE 8 21 344 447 OWN 43 12 162 217 RENT 103 9 29 411		1	02	23	
RENT 88 17 469 574 years 158 32 1068 1258 MORTGAGE 79 23 600 702 OWN 14 1 77 92 RENT 65 8 391 464 M/a 228 42 805 1075 MORTGAGE 8 21 344 447 OWN 43 12 162 217 RENT 103 9 29 411					
9 years 158 32 168 1258 MORTGAGE 79 23 600 702 OWN 14 1 77 92 RENT 65 8 391 464 n/s 28 42 805 1075 MORTGAGE 8 2 21 344 447 OWN 43 12 162 217 RENT 103 9 29 411					
MORTGAGE 79 23 600 702 OWN 14 1 77 92 RENT 65 8 391 464 n/a 28 42 805 1075 MORTGAGE 82 21 344 447 OWN 43 12 162 217 RENT 103 9 299 411					
OWN 14 1 77 92 RENT 65 8 391 464 N/a 228 42 805 1075 OWN 82 21 344 447 OWN 43 12 162 217 RENT 103 9 29 411					
RENT 65 8 391 464 n/a 228 42 805 1075 MORTGAGE 82 21 344 447 OWN 43 12 162 127 RENT 103 9 299 411					
v/s 228 42 805 1075 MORTGAGE 82 21 34 447 OWN 43 12 162 217 RENT 103 9 29 411					
MORTGAGE 82 21 344 447 OWN 43 12 162 217 RENT 103 9 299 411					
OWN 43 12 162 217 RENT 103 9 299 411					
RENT 103 9 299 411					
Grand Total 5627 1140 32950 39717					
	Grand Total	56	27 11	40 3	2950 39717

This is an example of the analysis conducted based on the years of experience of the person that has an impact on the loan status, considering home ownership also as a factor.

V. Observations & Conclusions

Based on the analysis conducted, below are the conclusions that we could come up with:

- It seems that **higher interest rates** (e.g., 10.00%, 10.25%, etc.) have a higher frequency of borrowers in the 'Charged Off' category compared to 'Fully Paid'.
- It appears that borrowers with **higher employment lengths** (e.g., >10 years) have a higher frequency in the 'Charged Off' category compared to 'Fully Paid'.
- It appears that borrowers who are **renting** (RENT) have a higher frequency in the 'Charged Off' category compared to 'Fully Paid', while borrowers with a mortgage (MORTGAGE) have a higher frequency in the 'Fully Paid' category.
- It seems that borrowers who took loans for purposes such as credit card, debt consolidation, and small business have higher frequencies in the 'Charged Off' category compared to 'Fully Paid'.
- So to conclude, it is safe to assume that the above factors need to be prioritized while approving a loan, like trying to find the home ownership status and the employment length of the person and then using that to understand the principal amount and interest rates that will be used.