

# General Analysis on Loan Data

Nicholas Wan

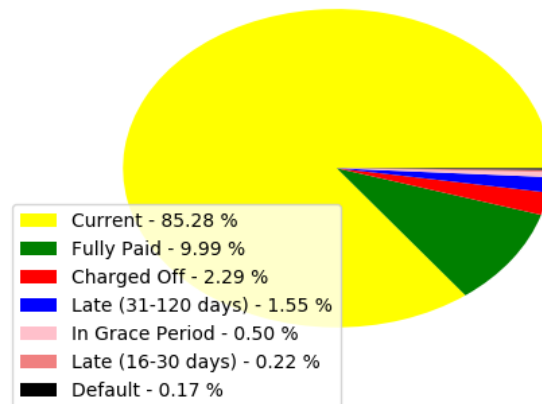
August 17, 2018

In this report, I will mainly discuss different features of 10000 granted loans in the past 5 years and whether these features have a positive or negative influence on the loans in terms of the status of the loans. In this experiment, we will be treating 10000 groups of loan data each consisting of 21 features and a status. By analyzing each feature, we will have a basic idea of the loans' performance.

## Status Overview :

There are 7 possible status for each loan. Good loans are the ones where the status is current or fully paid. Bad loans are the ones where the status is default, in grace period or late. This graph below shows the general overview all loans in terms of their status.

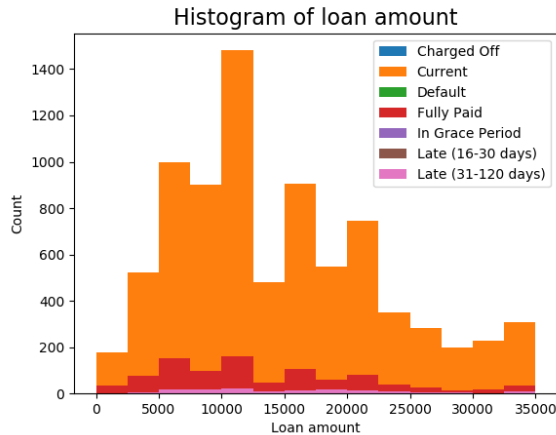
### Overall status of all small loans in the past 5 years



As we can see in this chart, good loans take up to 95.27% out of all loans granted. It is safe to say that the company is in good shape. Now, let's dive deeper into the data and see if we can gain more information about how to improve the performance of the loans.

## Loan Amount :

First of all, let's see if the amount of each loan influences its status:



In the histogram on the left, we can have a rough idea of the distribution of loan amount and the overall status among all loans. There are certain values that are more common than others. There are 720 loans with amount equal to 10000 and 304 loans with amount equal to 35000 .

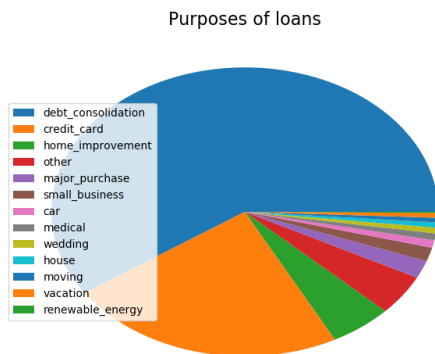
	Current	Fully Paid	Charged Off	In Grace Period	Sum	Good Loans	Bad Loans	Good Rate
0-5000	699	114	34	5	862.0	813	49.0	94.32 %
5000-10000	1899	251	49	11	2254.0	2150	104.0	95.39 %
10000-15000	1960	211	54	9	2268.0	2171	97.0	95.72 %
15000-20000	1452	165	41	10	1708.0	1617	91.0	94.67 %
20000-25000	1095	119	18	8	1268.0	1214	54.0	95.74 %
25000-30000	481	39	9	2	548.0	520	28.0	94.89 %
30000-35000	273	27	7	2	312.0	300	12.0	96.15 %
35000-40000	263	25	6	1	304.0	288	16.0	94.74 %

The table above shows us the detail of the status of loans with different amount. For the sake of page space, I hide 3 columns in this table(Default, Late(31-120 days) and Late(16-30 days)). You are welcome to examine the file *Loan\_Amount\_on\_Status\_full.csv* for full report.

In this table, each row indicates different loan amount interval represented by the left most column. The right most column indicates the good loan rate for that interval. We can see more clearly that most loans have the amount between 5,000 to 25,000 dollars. However, the good rate are similarly good among all intervals. As a result, for now we can conclude that the loan amount does not have major influence on loan status. Loans will not turn worse because of large amount or vice versa.

### Genral Purpose :

Now, we are going to analyze if the purpose of the loans have a major impact the loan status:



In the histogram on the left, we can have a rough idea of the distribution of general purpose of all the loans. As we can see, most of the loans are used for debt consolidation and credit card. These two purposes make up to 80% of all loans. As a result, the good performance on loans with these two purpose will be significant to the good overall performance.

	Current	Fully Paid	Charged Off	In Grace Period	Sum	Good Loans	Bad Loans	Good Rate
car	68	12	0	0	81.0	80	1.0	98.77 %
credit card	1957	180	30	7	2214.0	2137	77.0	96.52 %
debt consolidation	4829	556	142	32	5665.0	5385	280.0	95.06 %
home improvement	415	64	4	1	497.0	479	18.0	96.38 %
house	39	12	2	0	55.0	51	4.0	92.73 %
major purchase	153	28	2	1	189.0	181	8.0	95.77 %
medical	56	9	4	2	72.0	65	7.0	90.28 %
moving	39	9	1	0	50.0	48	2.0	96.00 %
other	350	53	16	3	431.0	403	28.0	93.50 %
renewable energy	10	1	1	0	12.0	11	1.0	91.67 %
small business	111	18	10	1	147.0	129	18.0	87.76 %
vacation	42	6	2	0	50.0	48	2.0	96.00 %
wedding	53	3	4	1	61.0	56	5.0	91.80 %

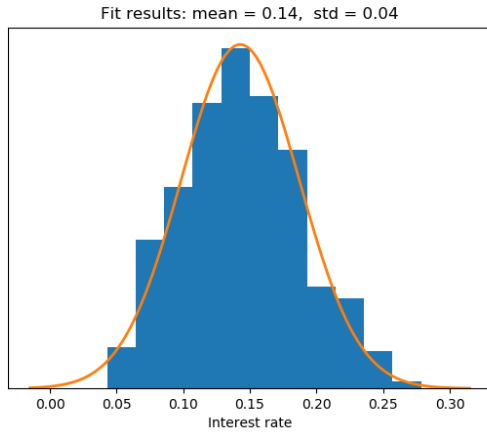
Now, let's get closer at each purpose to see their performance.

The table above shows us the detail of the status of loans with different purpose. For the sake of page space, I hide 3 columns in this table(Default, Late(31-120 days) and Late(16-30 days)). You are welcome to examine the file *Purpose\_on\_Status\_full.csv* for full report.

First let's go to the two types of loans that care the most about: debt consolidation and credit card. Fortunately, they are both performing well with 95.06% and 96.52% good rate. If we want to improve the overall status on this two types, I recommend improving the loans with purpose of debt consolidation because its rate still has increasing space and it is the most common purpose of all.

Also, there are loans with certain purposes require our attention. Loans for small business have the worst good loan rate while its quantity is not small. As a result, this is a major draw-back in the overall status. Being more carefully in granting loans for small business will be a good choice in the future. Finally, loans for medical, renewable energy and wedding also need more careful examination in the future although they have relatively small amount.

### Interest Rate :



In the histogram on the left, we can have a rough idea of the distribution of the interest rate of all the loans. As we can see, the interest rate are normally distributed and therefore we fit a normal distribution onto our data. Our distribution has  $mean = 0.14$  and standard deviation = 0.04. In general, our distribution is performing good in predicting interest rate for a single loan.

	Current	Fully Paid	Charged Off	In Grace Period	Sum	Good Loans	Bad Loans	Good Rate
0.0-0.02	0	0	0	0	0.0	0	0.0	0.00 %
0.02-0.04	0	0	0	0	0.0	0	0.0	0.00 %
0.04-0.06	0	0	0	0	0.0	0	0.0	0.00 %
0.06-0.08	842	144	16	2	1014.0	986	28.0	97.24 %
0.08-0.1	639	58	5	1	710.0	697	13.0	98.17 %
0.1-0.12	1071	105	8	7	1207.0	1176	31.0	97.43 %
0.12-0.14	1403	187	49	5	1664.0	1590	74.0	95.55 %
0.14-0.16	1545	181	38	10	1810.0	1726	84.0	95.36 %
0.16-0.18	1008	109	31	7	1184.0	1117	67.0	94.34 %
0.18-0.2	772	82	35	7	921.0	854	67.0	92.73 %
0.2-0.22	417	39	20	5	496.0	456	40.0	91.94 %
0.22-0.24	297	33	14	3	365.0	330	35.0	90.41 %
0.24-0.26	123	13	2	1	147.0	136	11.0	92.52 %
0.26-0.28	5	0	0	0	6.0	5	1.0	83.32 %
0.28-0.3	0	0	0	0	0.0	0	0.0	0.00 %

Furthermore, we are going to analyze the performance of loans in different interest rate interval according to table above. For the sake of page space, I hide 3 columns in this table(Default, Late(31-120 days) and Late(16-30 days)). You are welcome to examine the file *Interest\_Rate\_on\_Status\_full.csv* for full report. In this table, different rows represent different interest rate interval indicated by the left most column and the right most column tells us the good loan rate for that interval. Starting from 0.06 up to 0.28, we can observe that no whether the amount of loans in the intervals are small or big, the good loan rate is constantly decreasing.

As a result, I believe it is safe to conclude that the higher the interest rate of one loan is, the bigger risk there is for that loan to turn bad.

#### Conclusion :

Based on my evidence and analysis above, there are several conclusions that I can safely draw from the data sample:

- I) The status of all loans is generally good and good loans(current of fully paid) make up 95.27% of all loans;
- II) Loan amount do not have major impact on loan status. Loans have stable performance with small or large amount;
- III) Loans that are used for debt consolidation make up the majority of all loans and loans that are used for debt consolidation or credit card make up to over 80% of all loans;
- IV) Loans that are used for small business perform the worst among all loans;
- V) Loans that used for renewable energy, medical or wedding perform poorly but with fairly small amount;
- VI) The interest rate of all loans are normally distributed with approximately  $mean = 14\%$  and  $std = 0.04$ ;
- VII) The higher the interest rate is, the worse the loan is likely to perform.