

Lending Club Case Study

Teja Rajeev Tenneti

Objective and Dataset Analysis

- Objective of the case study is to identify patterns that a person is likely to default on a loan from the loan.csv dataset that has been provided.
- Dataset contains 39717 rows and 111 columns .
- 54 Columns are completely NULL , 3 columns have more than 50% Null values, upto 7 columns dropped which have same values through out .
- Only keeps rows where loan_status != 'Current'
- 85% of the customers are Fully Paid , and 14.5% of Customers are Charged-Off

- Observations from Correlation Matrix
- As interest rate and term increase the loan defaults increase
- Annual income increases defaulting decreases , weak -ve core with open_acc

```
: new_num_df.corr()
```

	loan_amnt	funded_amnt	funded_amnt_inv	term	int_rate	installment	emp_length	annual_inc	dti	open_acc	pub_rec_bankruptcies
loan_amnt	1.000000	0.981940	0.936901	0.349595	0.291145	0.931488	0.157613	0.403358	0.090237	0.166060	-0.02894
funded_amnt	0.981940	1.000000	0.954829	0.328801	0.296134	0.957120	0.157282	0.399203	0.089438	0.164063	-0.02999
funded_amnt_inv	0.936901	0.954829	1.000000	0.348418	0.288743	0.902616	0.166206	0.382540	0.097037	0.151554	-0.03054
term	0.349595	0.328801	0.348418	1.000000	0.435550	0.091692	0.105847	0.079962	0.078553	0.043168	0.01885
int_rate	0.291145	0.296134	0.288743	0.435550	1.000000	0.269155	0.009038	0.053952	0.111587	-0.008005	0.08812
installment	0.931488	0.957120	0.902616	0.091692	0.269155	1.000000	0.131165	0.398460	0.079374	0.160841	-0.02690
emp_length	0.157613	0.157282	0.166206	0.105847	0.009038	0.131165	1.000000	0.212148	0.058753	0.099203	0.05254
annual_inc	0.403358	0.399203	0.382540	0.079962	0.053952	0.398460	0.212148	1.000000	-0.072526	0.265014	-0.00510
dti	0.090237	0.089438	0.097037	0.078553	0.111587	0.079374	0.058753	-0.072526	1.000000	0.298389	0.00525
open_acc	0.166060	0.164063	0.151554	0.043168	-0.008005	0.160841	0.099203	0.265014	0.298389	1.000000	0.01010
bankruptcies	-0.028946	-0.029997	-0.030540	0.018853	0.088120	-0.026909	0.052544	-0.005109	0.005256	0.010105	1.00000
loan_status	0.066779	0.064207	0.044905	0.176145	0.212308	0.033540	0.009259	-0.066396	0.041266	-0.006908	0.04757

Univariate Analysis

- Mean loan amount is 10681
- Mean Annual Income is 60,829
- Average Employment length is 4.5 Years

```
] new_num_df.describe()
```

	loan_amnt	funded_amnt	funded_amnt_inv	term	int_rate	installment	emp_length	annual_inc	dti	open_acc
count	36642.000000	36642.000000	36642.000000	36642.000000	36642.000000	36642.000000	36642.000000	36642.000000	36642.000000	36642.000000
mean	10681.045658	10435.008324	9892.999247	41.847716	11.887884	311.863566	4.764942	60829.831120	13.447256	9.166366
std	7047.619569	6799.747627	6730.054355	10.303024	3.667161	199.215000	3.588478	27717.831464	6.661561	4.352874
min	500.000000	500.000000	0.000000	36.000000	5.420000	15.690000	0.000000	4000.000000	0.000000	2.000000
25%	5000.000000	5000.000000	4975.000000	36.000000	8.940000	163.080000	2.000000	40000.000000	8.350000	6.000000
50%	9250.000000	9000.000000	8375.000000	36.000000	11.710000	270.410000	4.000000	56000.000000	13.580000	8.000000
75%	15000.000000	14400.000000	13618.409453	36.000000	14.270000	408.480000	8.000000	77704.050000	18.717500	12.000000
max	35000.000000	35000.000000	35000.000000	60.000000	24.400000	1305.190000	10.000000	140000.000000	29.990000	44.000000

Univariate Analysis

- Maximum number of customers are B grade credit rating >10,000
- Majority of home ownership of customers who took loans stay in rented homes
- Maximum customers do not have their income from verified sources >16000