

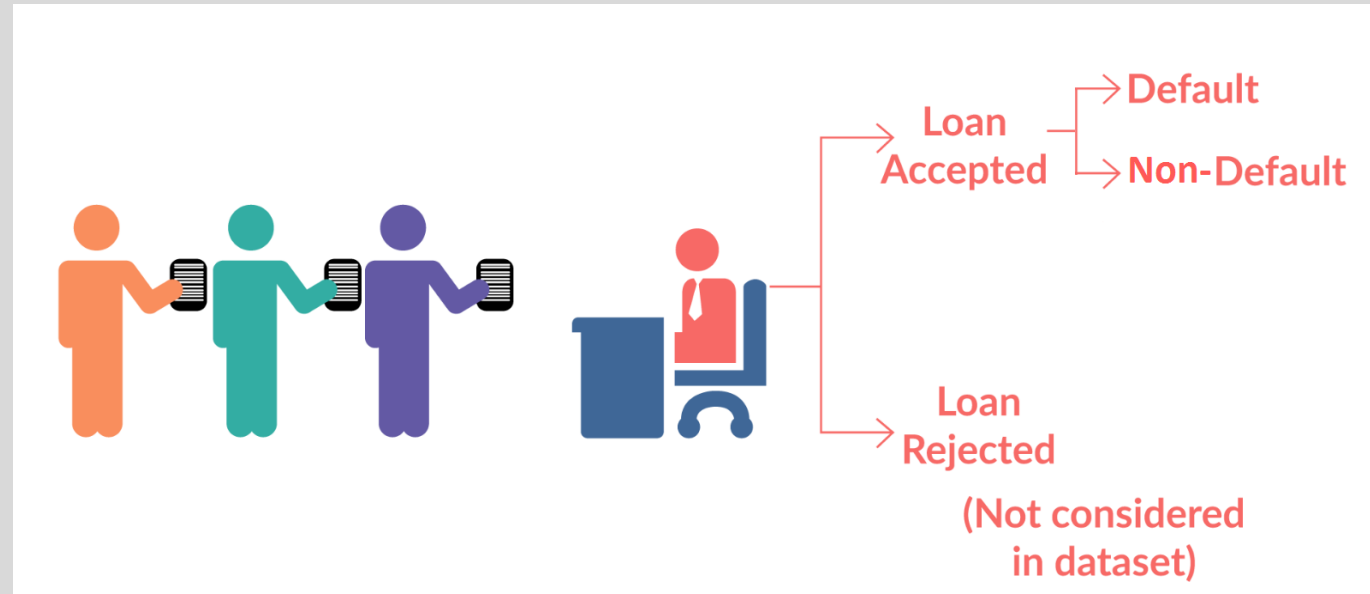
Lending Club Case Study

Ashish Singh
Roopali Paliwal

Problem Statement

You work for a **consumer finance company** which specializes in lending various types of loans to urban customers. When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile. Two **types of risks** are associated with the bank's decision:

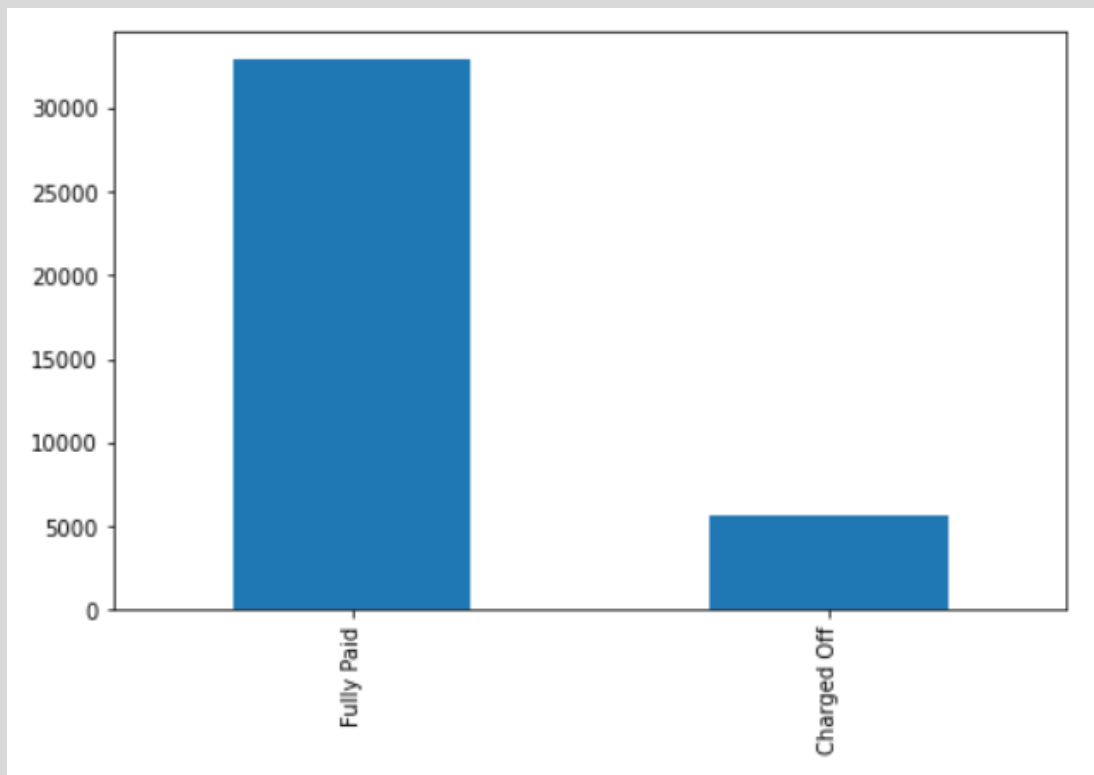
- If the applicant is **likely to repay the loan**, then not approving the loan results in a **loss of business** to the company
 - If the applicant is **not likely to repay the loan**, i.e. he/she is likely to default, then approving the loan may lead to a **financial loss** for the company
- In this case study, we will use EDA to understand how **consumer attributes** and **loan attributes** influence the tendency of default.



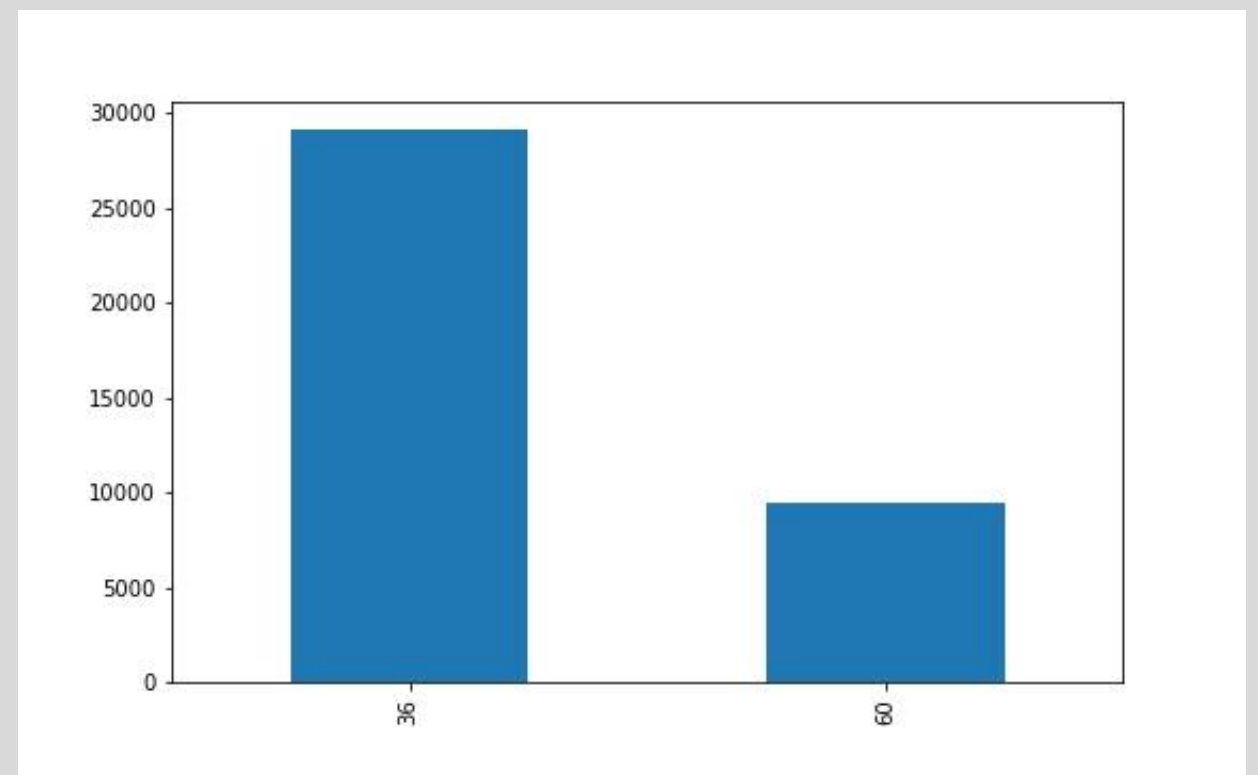
Univariate Analysis Observation

Following are the observations on loan data set after cleaning process

Fully Paid vs Charged off customers

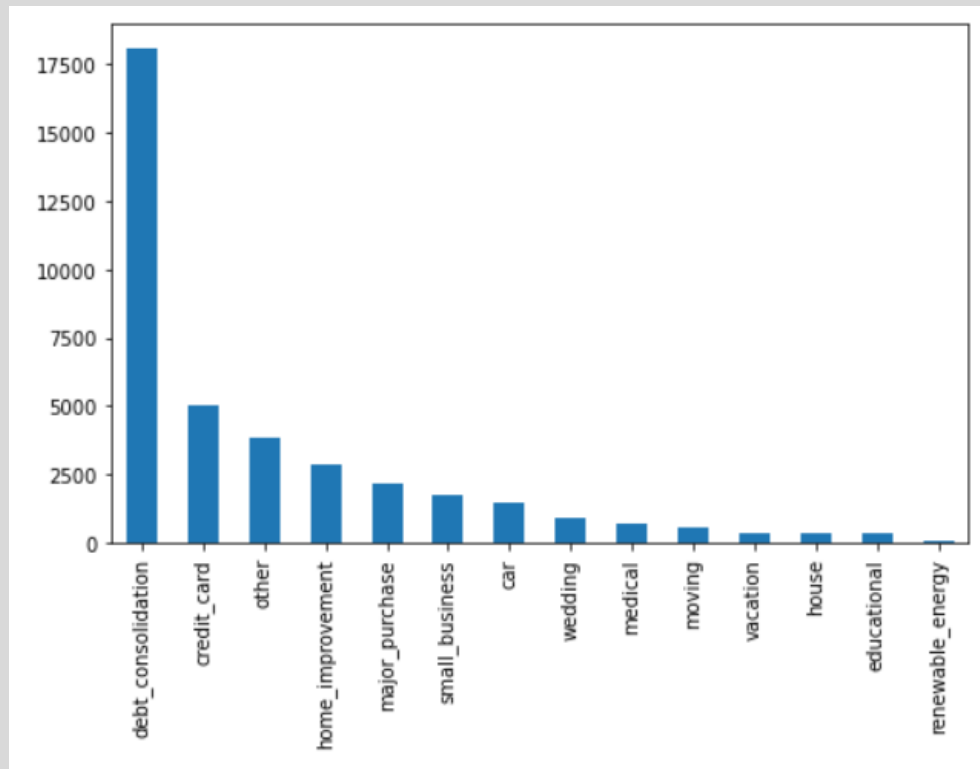


36 vs 60 month loans

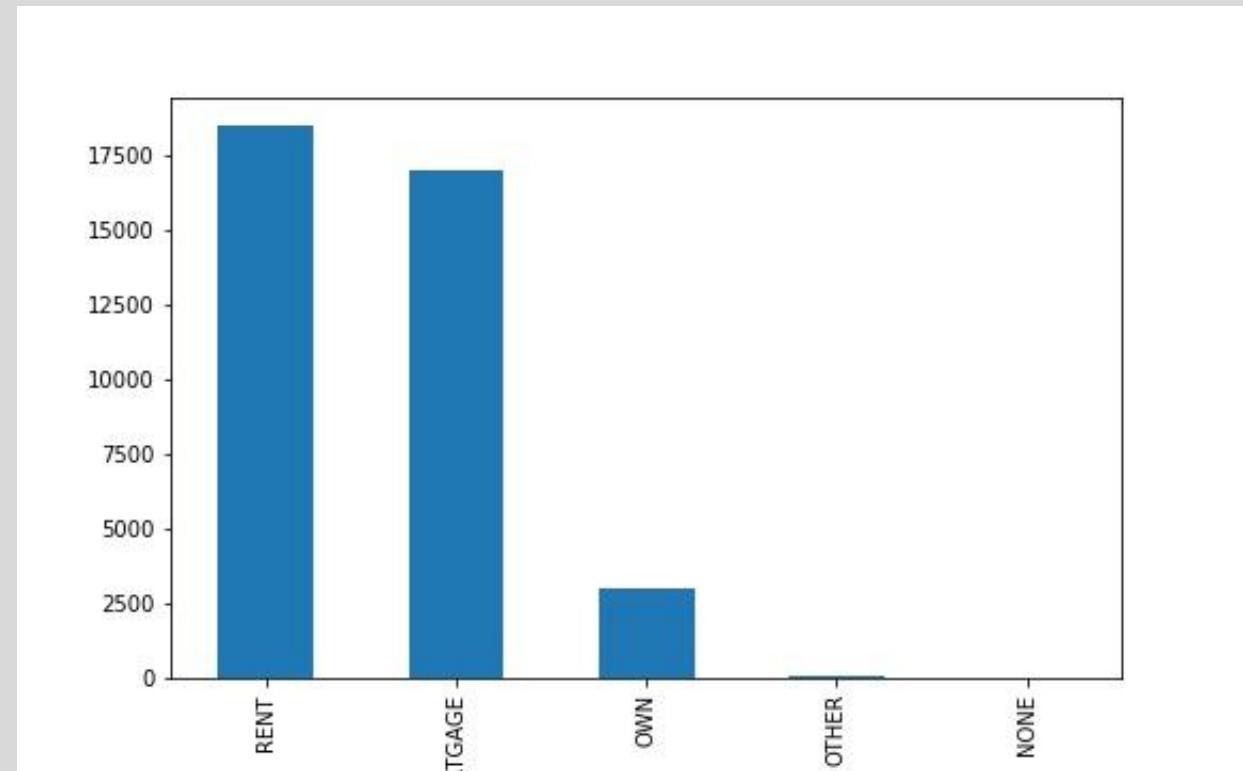


Univariate Analysis Observation

Purpose of loan

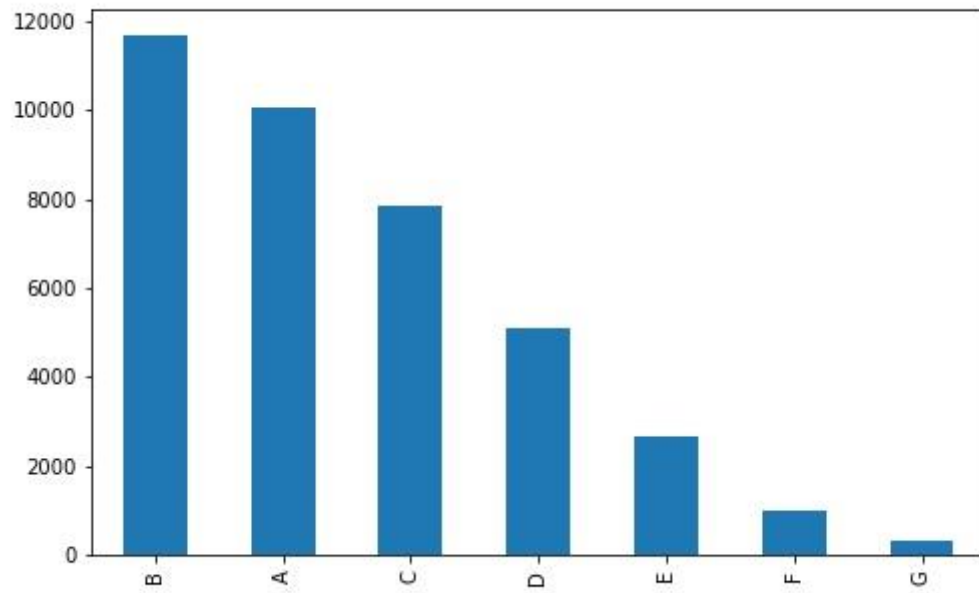


House Ownership Data

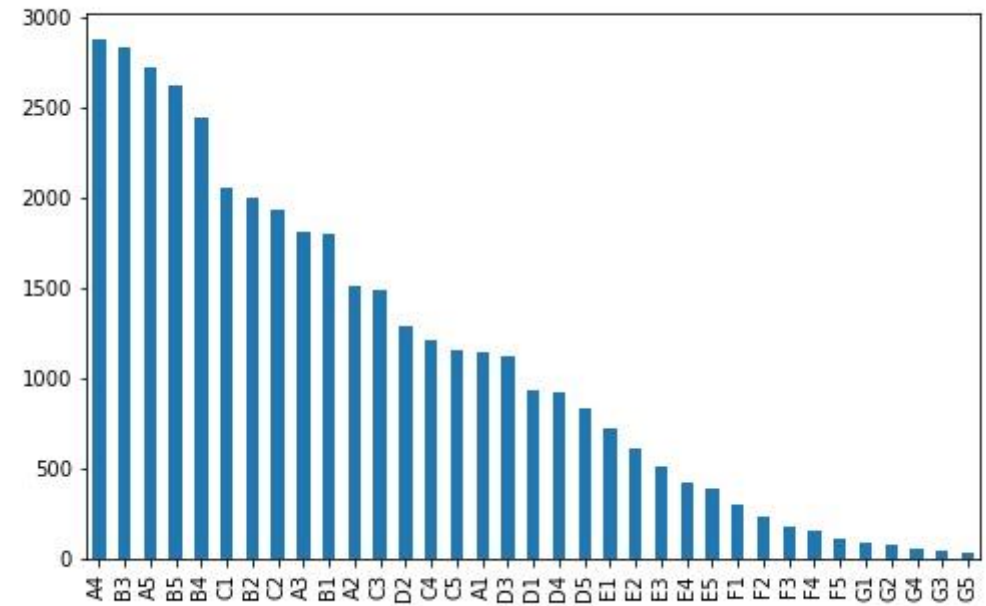


Univariate Analysis Observation

Frequency for Grades



Frequency for Sub-Grades

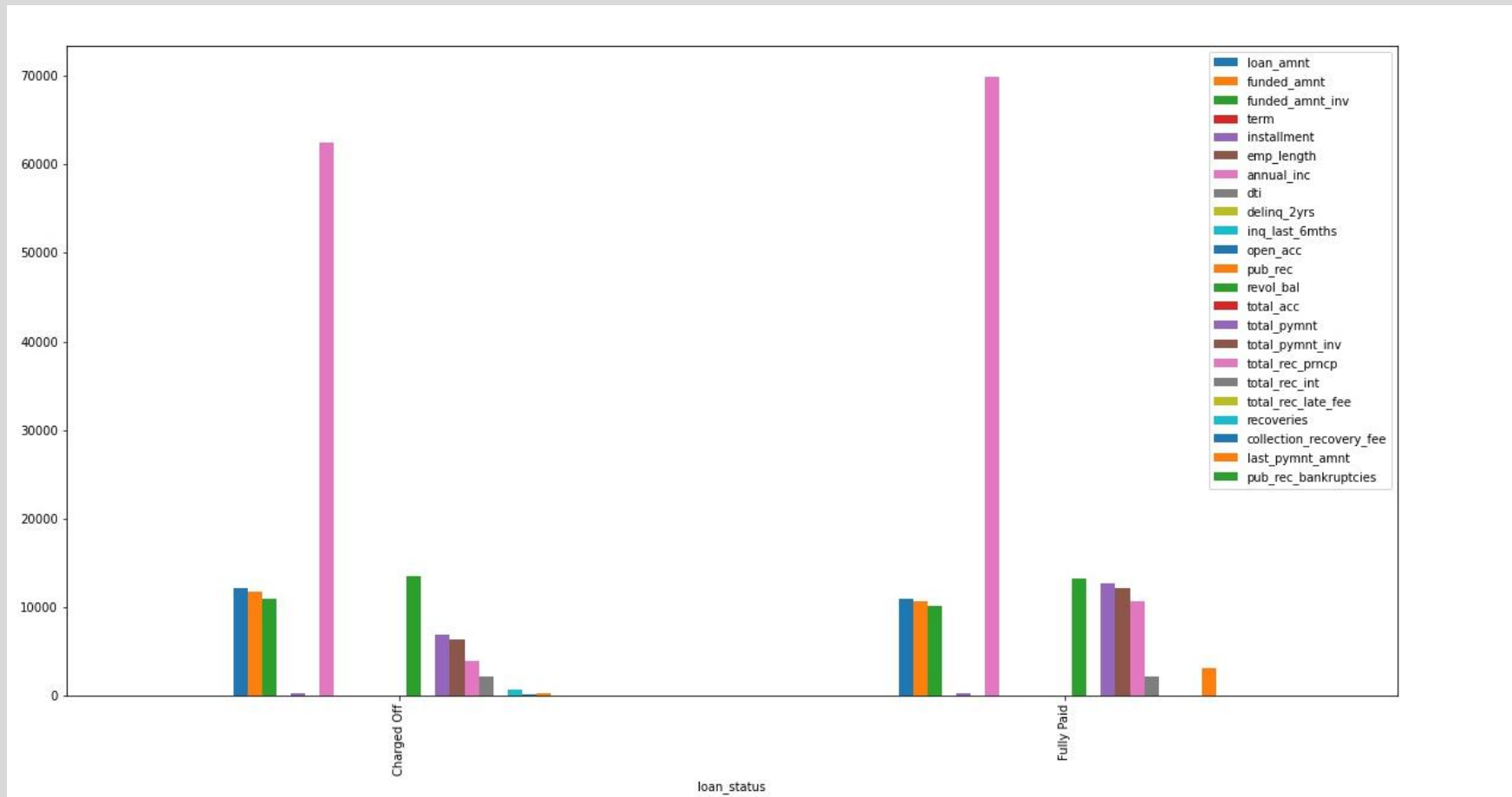


Univariate Analysis Observation

	Col	Mean	Median	Mode
0	loan_amnt	11047.025430	9600.000000	0 10000 dtype: int64
1	funded_amnt	10784.058506	9550.000000	0 10000 dtype: int64
2	funded_amnt_inv	10222.481123	8733.440000	0 5000.0 dtype: float64
3	term	41.898437	36.000000	0 36 dtype: int64
4	installment	322.466318	277.860000	0 311.11 dtype: float64
5	emp_length	5.061048	4.000000	0 10.0 dtype: float64
6	annual_inc	68777.973681	58868.000000	0 60000.0 dtype: float64
7	dti	13.272727	13.370000	0 0.0 dtype: float64
8	delinq_2yrs	0.146668	0.000000	0 0 dtype: int64
9	inq_last_6mths	0.871737	1.000000	0 0 dtype: int64
10	open_acc	9.275423	9.000000	0 7 dtype: int64
11	pub_rec	0.055422	0.000000	0 0 dtype: int64
12	revol_bal	13289.489826	8762.000000	0 0 dtype: int64
13	total_acc	22.052648	20.000000	0 16 dtype: int64
14	total_pymnt	11866.970365	9674.048405	0 11196.56943 dtype: float64
15	total_pymnt_inv	11271.196153	9066.660000	0 0.0 dtype: float64
16	total_rec_prncp	9648.322898	8000.000000	0 10000.0 dtype: float64
17	total_rec_int	2119.243370	1300.450000	0 0.0 dtype: float64
18	total_rec_late_fee	1.368634	0.000000	0 0.0 dtype: float64
19	recoveries	98.035545	0.000000	0 0.0 dtype: float64
20	collection_recovery_fee	12.772728	0.000000	0 0.0 dtype: float64
21	last_pymnt_amnt	2746.243218	568.260000	0 0.0 dtype: float64
22	pub_rec_bankruptcies	0.043479	0.000000	0 0.0 dtype: float64

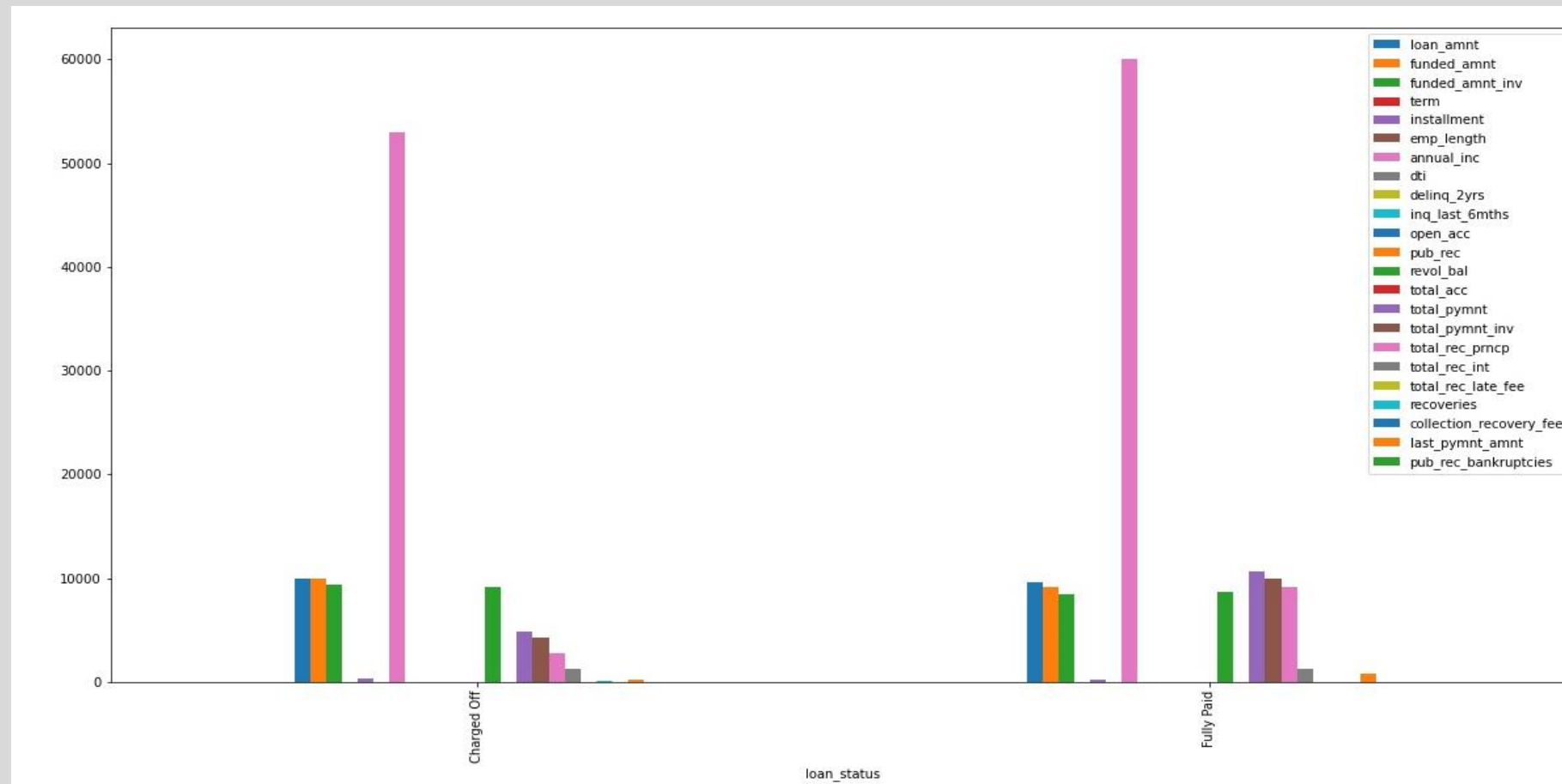
Segmented Univariate Analysis Observation

Mean distribution by loan_status



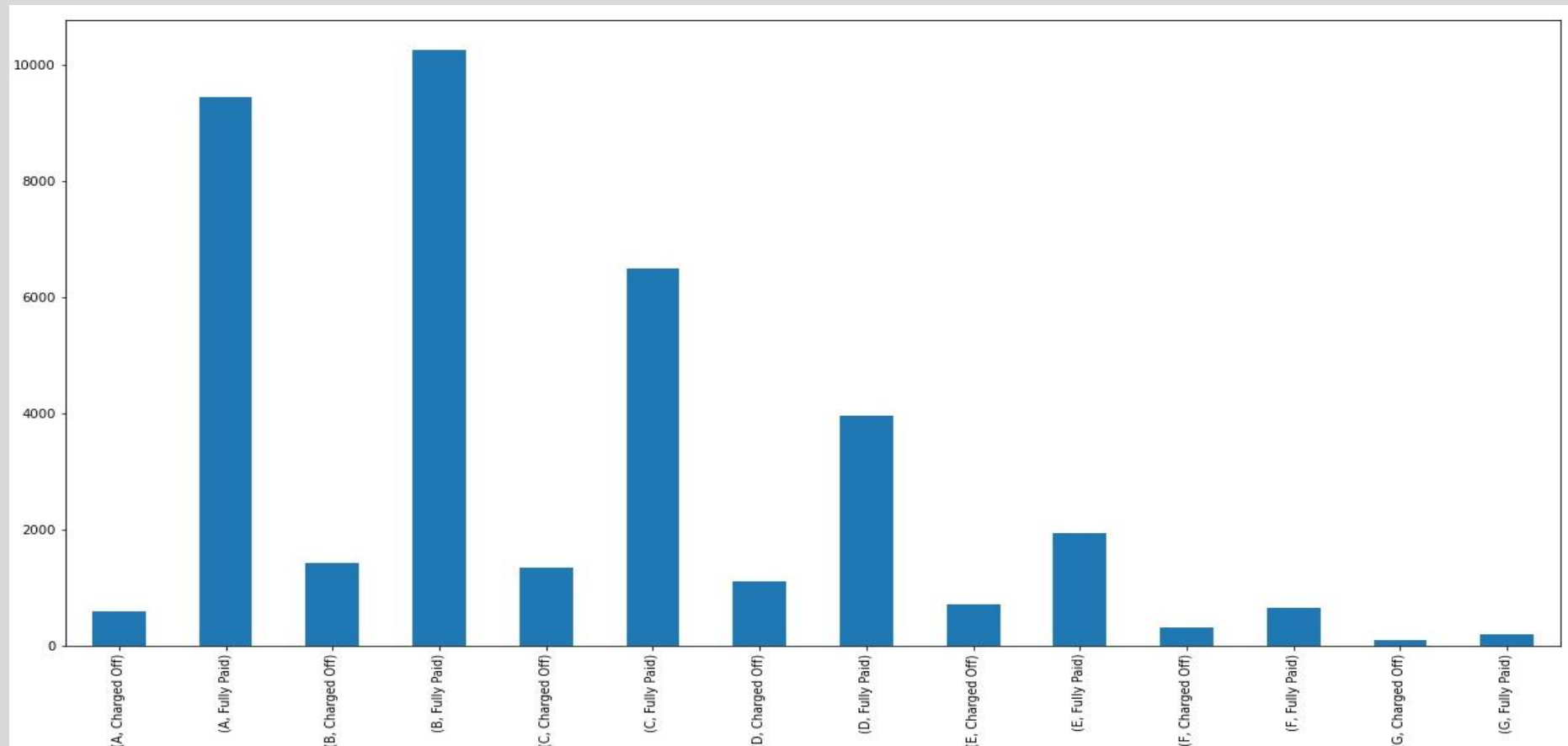
Segmented Univariate Analysis Observation

Median distribution by loan_status



Segmented Univariate Analysis Observation

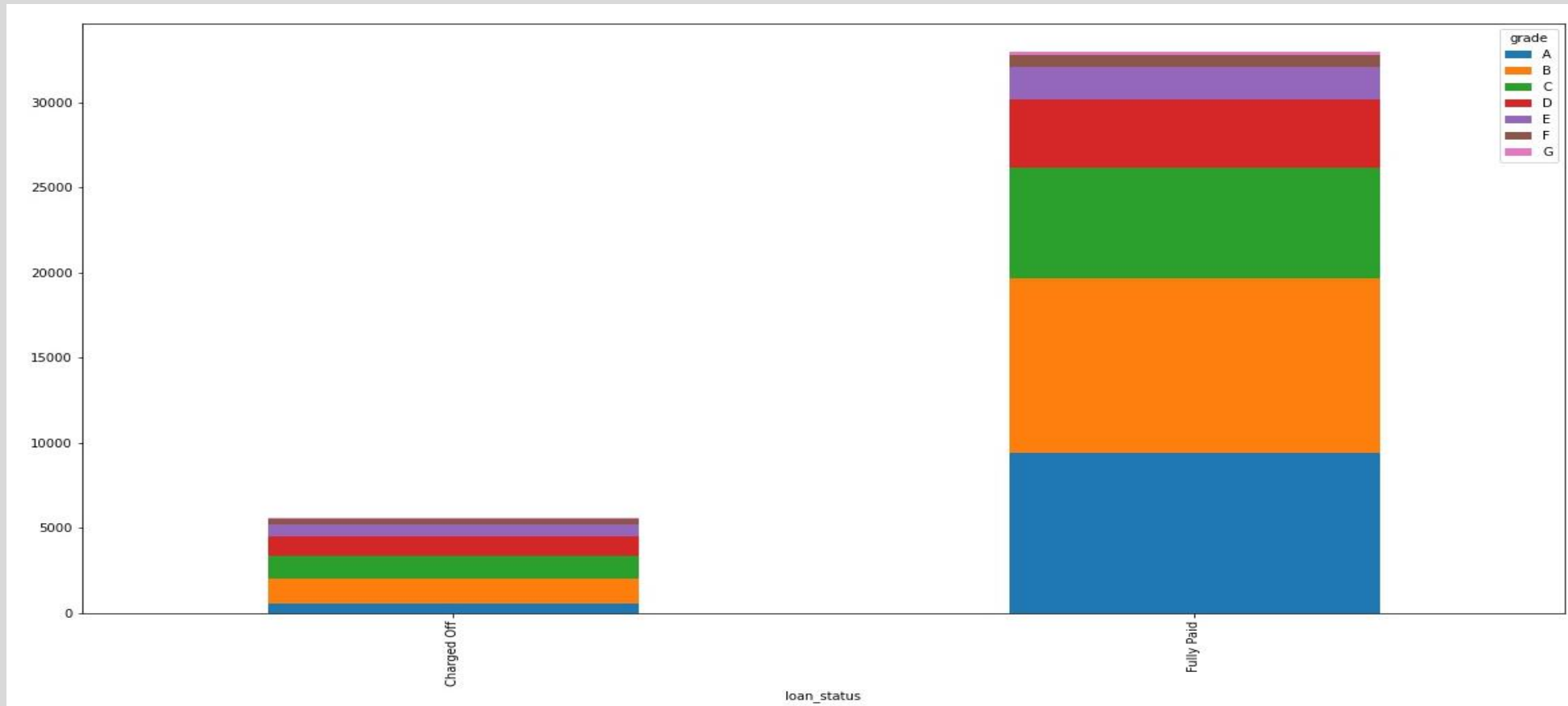
Observation - People with grade A, B and C have higher ratio of fully paid loans which decreases as the grade increase.



Distribution of data based on Grades and loan_status

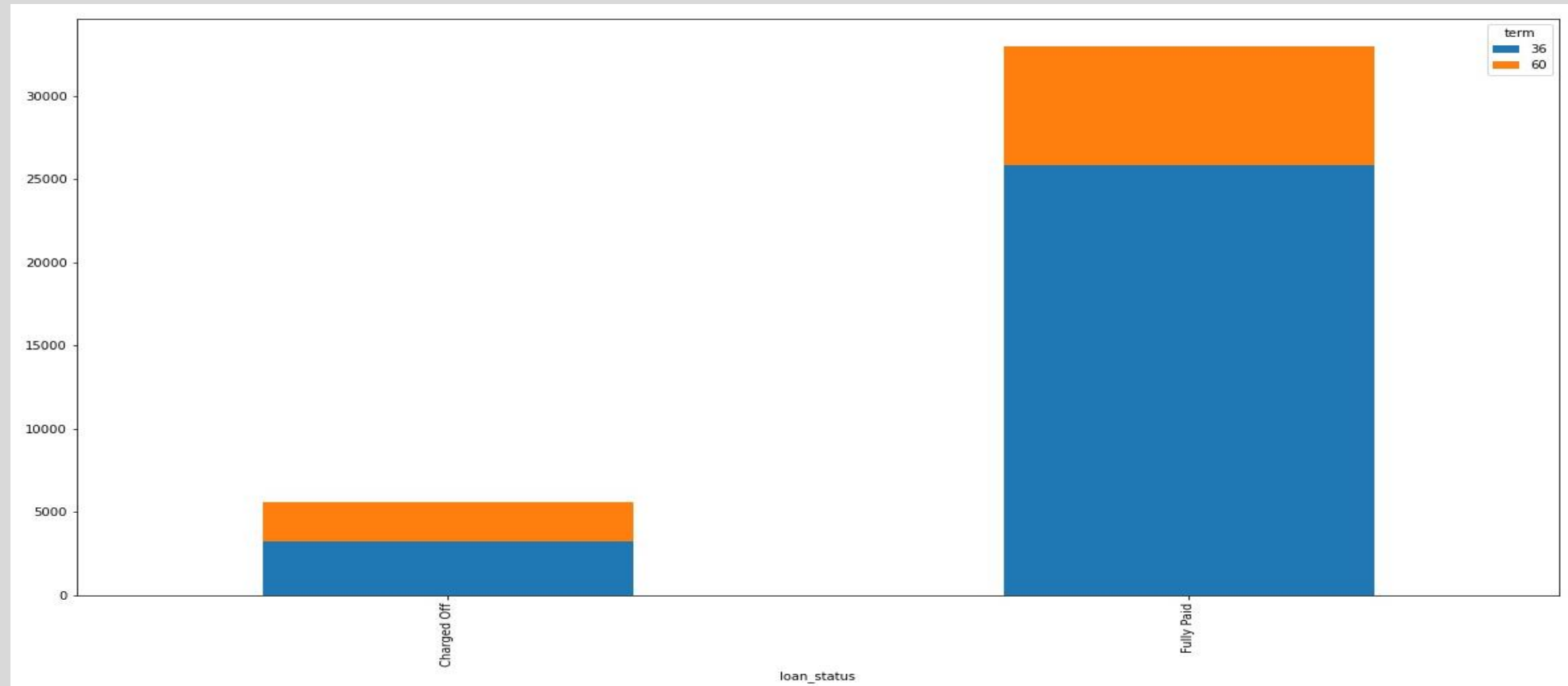
Segmented Univariate Analysis Observation

Another representation of ratio of grade by loan_status



Segmented Univariate Analysis Observation

Loans with tenure of 36 months have a higher ratio of being fully paid than 60 months.



Distribution of data based on Term and loan_status

Bivariate Analysis: Correlation

Correlation between different continuous variables

	installment	emp_length	annual_inc	dti	delinq_2yrs	inq_last_6mths	open_acc	pub_rec	revol_bal	total_acc	total_pymnt_inv	recoveries	collection_recovery_fee	last_pymnt_amnt	pub_rec_bankruptcies
installment	1	0.121101024	0.267841852	0.052038485	-0.019755423	0.011014363	0.172893482	-0.045705537	0.309500943	0.229859671	0.817664953	0.121463394	0.077518921	0.413588121	-0.033360603
emp_length	0.121101024	1	0.111107342	0.048619424	0.015420207	0.010509461	0.096526506	0.064306088	0.152201545	0.208554941	0.141445669	0.026485332	0.007130474	0.083555051	0.063211525
annual_inc	0.267841852	0.111107342	1	-0.121529558	0.02222892	0.035465053	0.156926895	-0.017863963	0.277374436	0.234534414	0.245198424	0.02218392	0.01598135	0.143242312	-0.016223526
dti	0.052038485	0.048619424	-0.121529558	1	-0.033346952	0.002178367	0.287849024	-0.004742436	0.228090104	0.22911882	0.066276797	0.02616237	0.011743569	0.008566188	0.005966447
delinq_2yrs	-0.019755423	0.015420207	0.02222892	-0.033346952	1	0.008733414	0.011648619	0.007622995	-0.054437252	0.06699432	-0.029608041	0.012442522	0.014077215	-0.012511172	0.003404181
inq_last_6mths	0.011014363	0.010509461	0.035465053	0.002178367	0.008733414	1	0.093434304	0.023726276	-0.020787001	0.113516075	-0.018588948	0.018899169	0.012387849	0.027772392	0.01551708
open_acc	0.172893482	0.096526506	0.156926895	0.287849024	0.011648619	0.093434304	1	2.83629E-05	0.288054554	0.687259916	0.152007542	0.017241314	0.006674176	0.082396357	0.005803161
pub_rec	-0.045705537	0.064306088	-0.017863963	-0.004742436	0.007622995	0.023726276	2.83629E-05	1	-0.060502262	-0.023494222	-0.053077228	-0.005799676	-0.005731071	-0.033439553	0.845938454
revol_bal	0.309500943	0.152201545	0.277374436	0.228090104	-0.054437252	-0.020787001	0.288054554	-0.060502262	1	0.312633772	0.273032554	0.043593026	0.023254022	0.125315302	-0.048572244
total_acc	0.229859671	0.208554941	0.234534414	0.22911882	0.06699432	0.113516075	0.687259916	-0.023494222	0.312633772	1	0.219402808	0.024023916	0.010952154	0.166888109	-0.011060183
total_pymnt_inv	0.817664953	0.141445669	0.245198424	0.066276797	-0.029608041	-0.018588948	0.152007542	-0.053077228	0.273032554	0.219402808	1	0.023405835	0.018541847	0.500067066	-0.046435691
recoveries	0.121463394	0.026485332	0.02218392	0.02616237	0.012442522	0.018899169	0.017241314	-0.005799676	0.043593026	0.024023916	0.023405835	1	0.796782373	-0.072355671	-0.004866462
collection_recovery_fee	0.077518921	0.007130474	0.01598135	0.011743569	0.014077215	0.012387849	0.006674176	-0.005731071	0.023254022	0.010952154	0.018541847	0.796782373	1	-0.043313861	-0.005369578
last_pymnt_amnt	0.413588121	0.083555051	0.143242312	0.008566188	-0.012511172	0.027772392	0.082396357	-0.033439553	0.125315302	0.166888109	0.500067066	-0.072355671	-0.043313861	1	-0.023483373
pub_rec_bankruptcies	-0.033360603	0.063211525	-0.016223526	0.005966447	0.003404181	0.01551708	0.005803161	0.845938454	-0.048572244	-0.011060183	-0.046435691	-0.004866462	-0.005369578	-0.023483373	1

Note:-

Shades of red demonstrate higher correlation

Shades of blue demonstrate lower correlation

Conclusion

The parameters **Term** and **Grade** seem to influence loan status.

- Loans with higher Grades (A,B) seem to have higher ratio of fully paid off
- Loans with Term of 36 months seem to have higher ratio of fully paid off.
- People with higher number of derogatory public records also seem to have higher number of bankruptcy record