Lending Club Case Study

SAI BHARATH PARSAM

UDAY RANGTA

Problem statement

The Lending Club company which is online platform for proving the loans wants to know which customers are likely in not repaying the loans based on information provided by them in application form

To identify the driving factors for loan default so that for those applicants , loan can be rejected which will save the company's investment and better business

Approach

1. Understand the problem statement

2. Do market research

3. Understand the given data attributes

4. Clean the data – all null value columns and partially null columns

5. Identify behavioral , categorical , loan and geography related attributes

6. Change the datatypes of identified attributes.

7. Remove behavioral attributes.

8. Create derived columns and bucket the required columns in to groups for better analysis

9. Univariate analysis

10. Segmented univariate analysis

11. Bivariate analysis

12. Conclusion

Identified variables

1. id
2. member_id
3. loan_amnt
4. funded_amnt
5. funded_amnt_inv
6. term
7. int_rate
8. installment
9. grade
10. sub_grade
11. emp_length
12. home_ownership
13. annual_inc
14. verification_status
15. issue_d
16. loan_status
17. purpose
18. addr_state
19. dti
20. pub_rec_bankruptcies

Overall, most of the loans are being fully paid. 15% defaults.

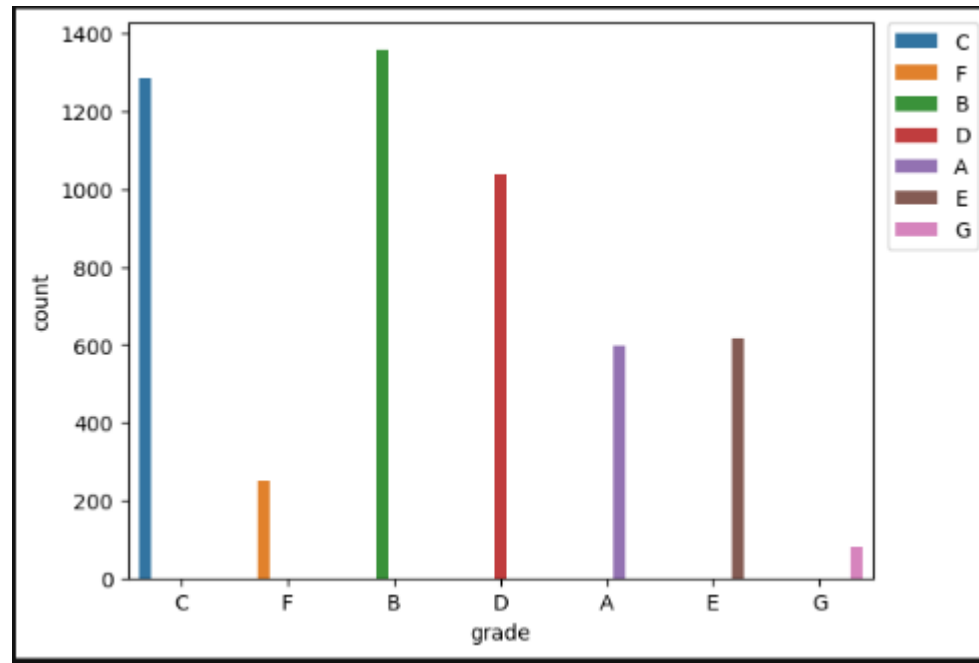# Over the years, default loan cases increased
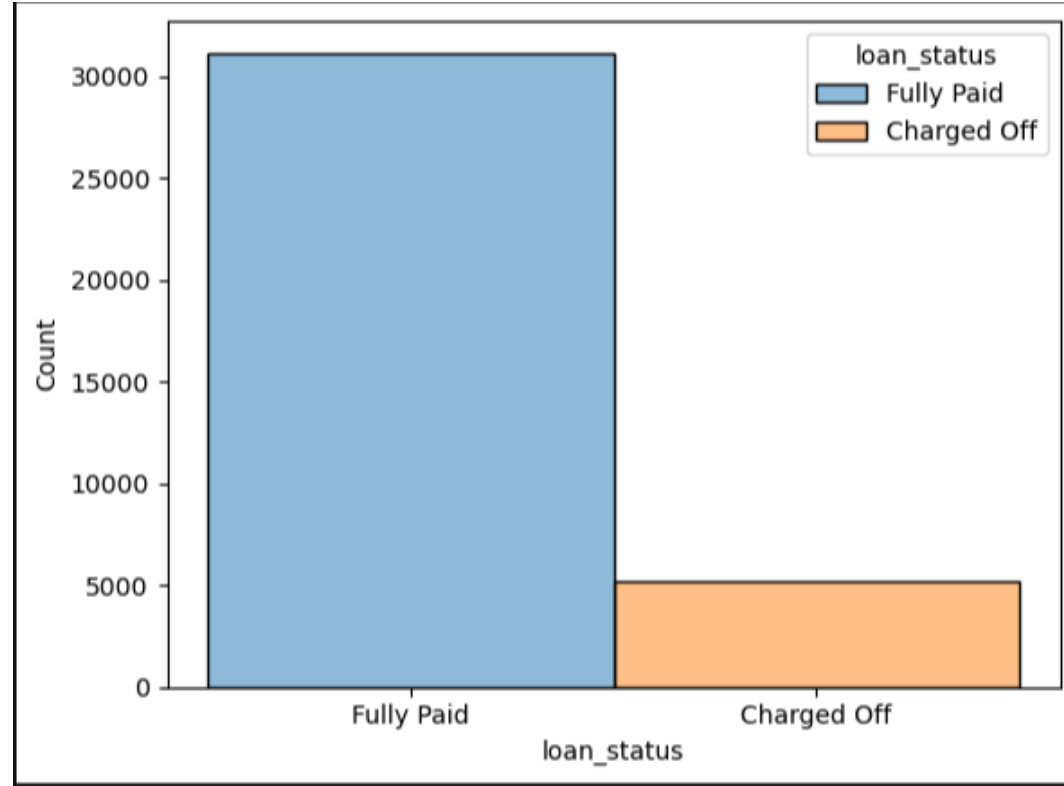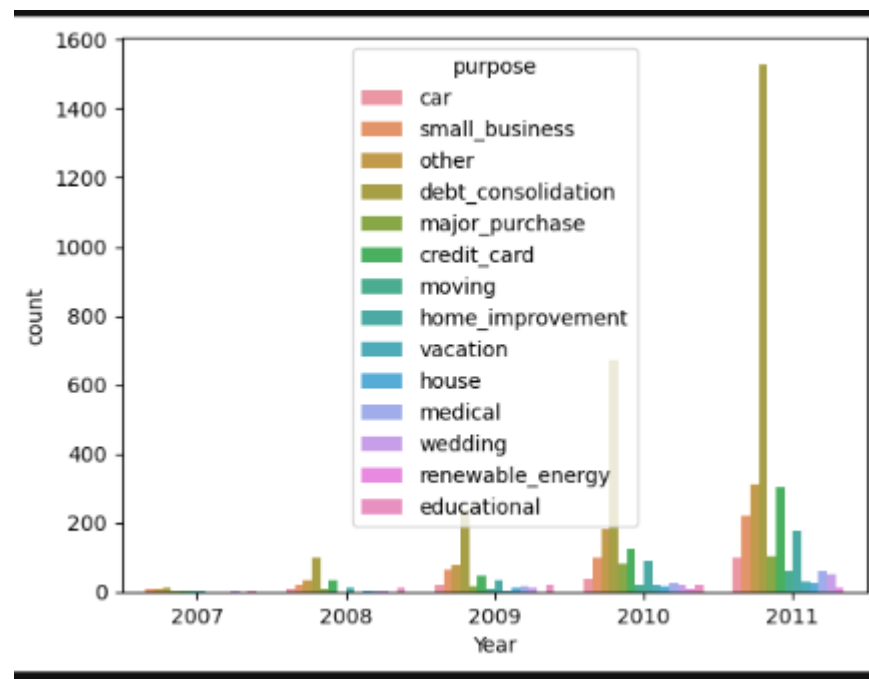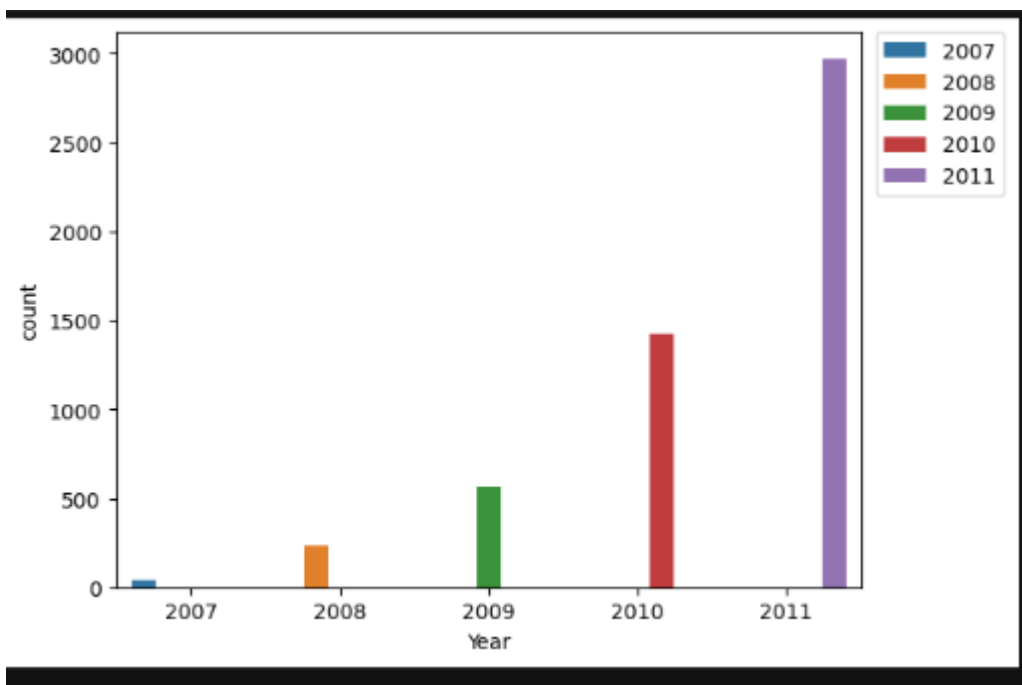
Rent and Mortgage home ownership has higher default rates
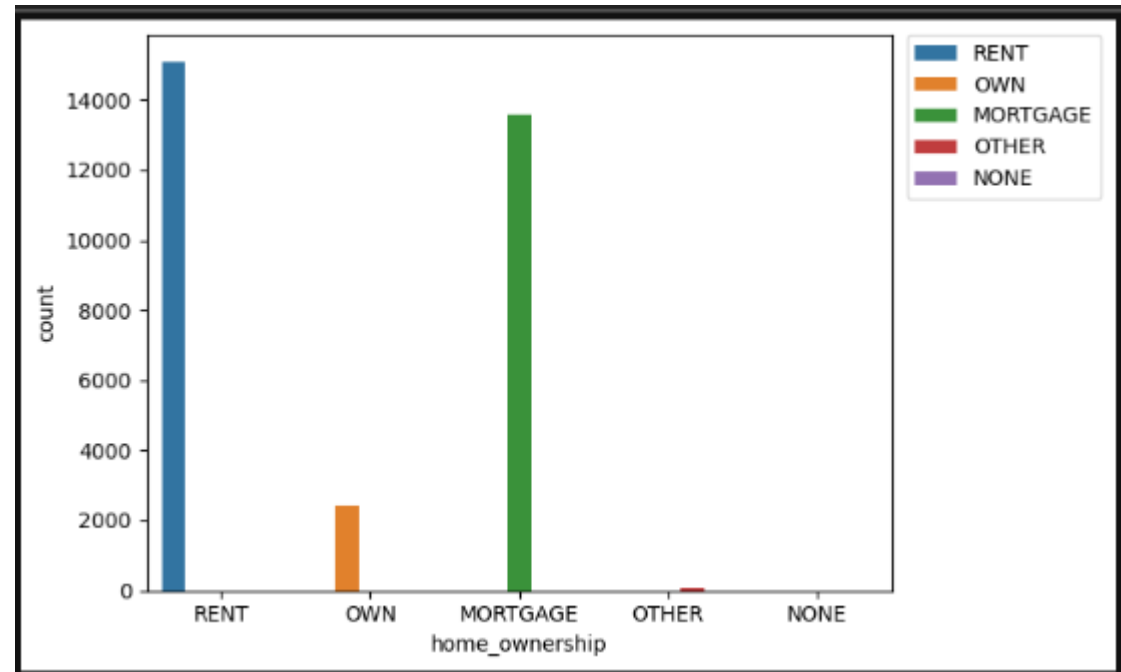
Higher the grade, higher the chance of defaults.

Medium and high loan amounts are prone to defaults.
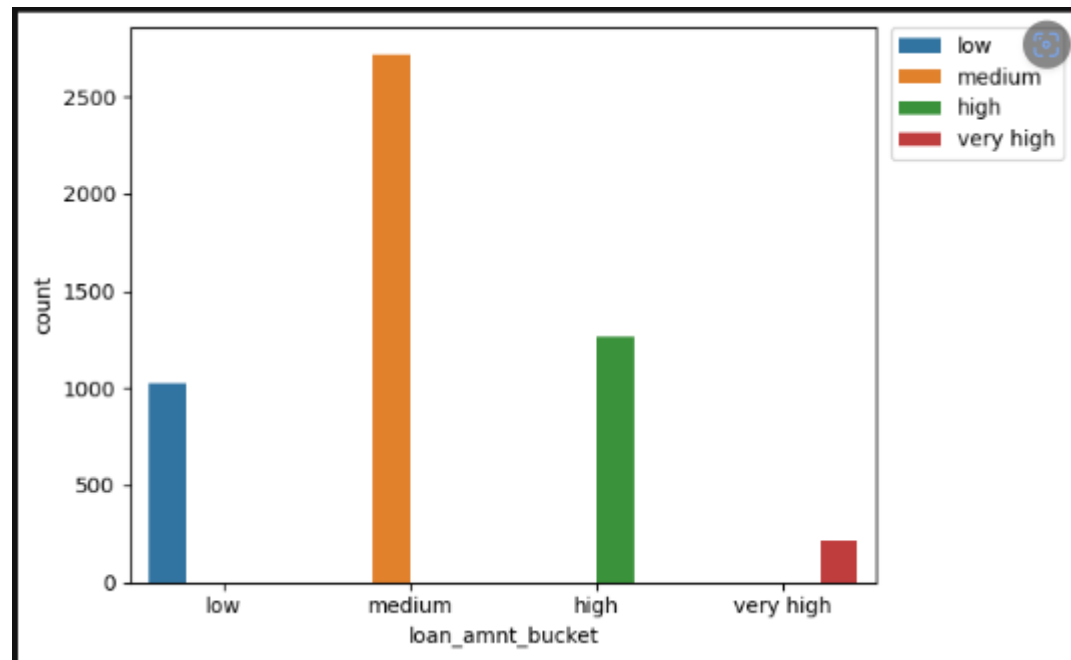
Higher the installments amount, higher the chance of defaults.
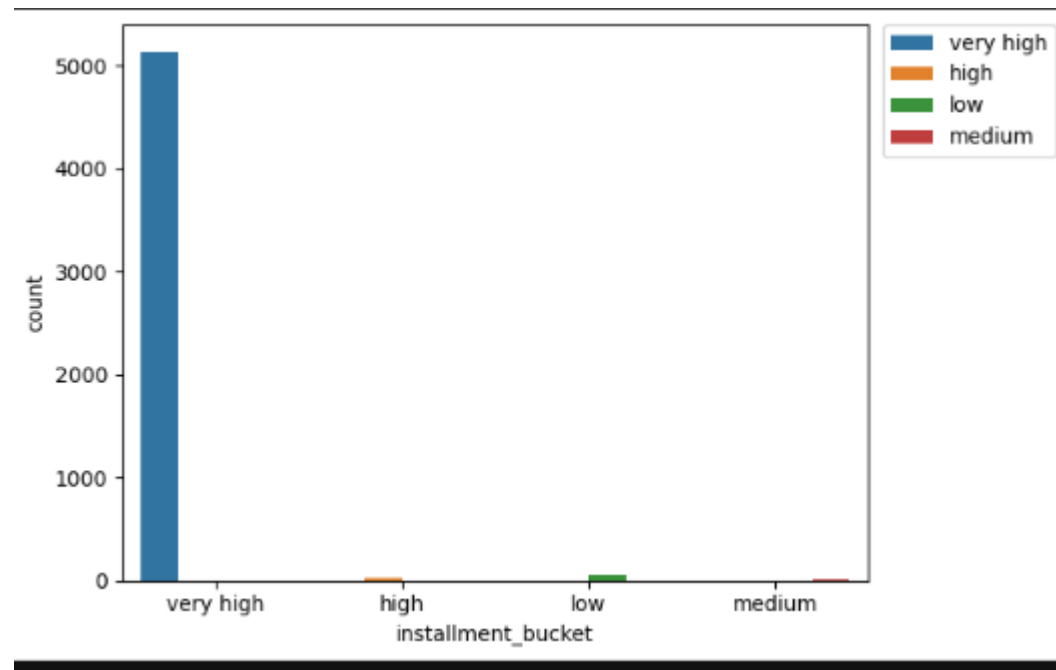
36 months term has high default rates
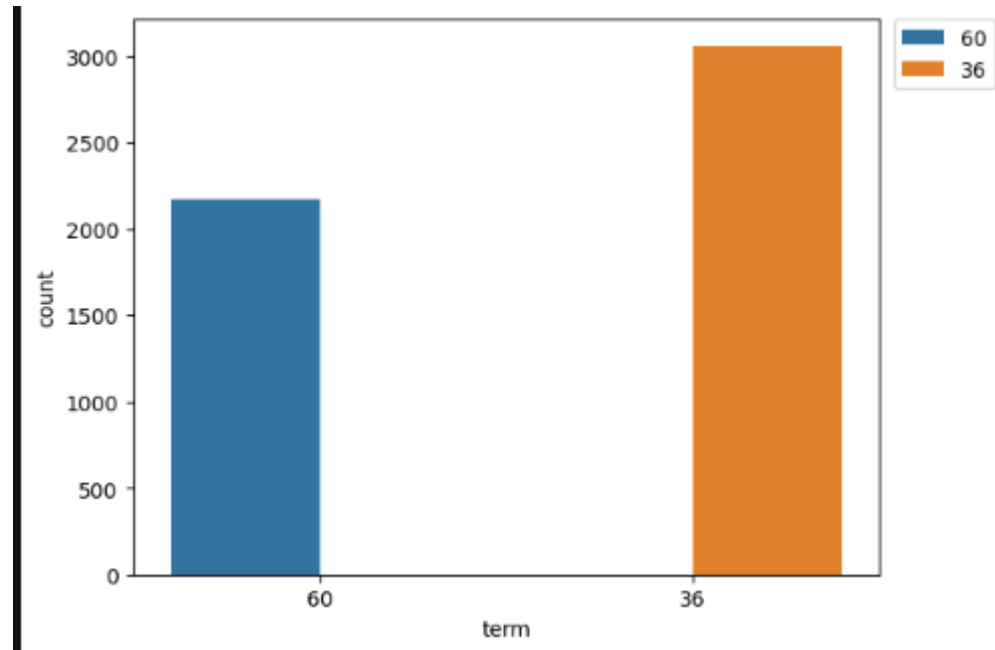
Low income earners are likely to default
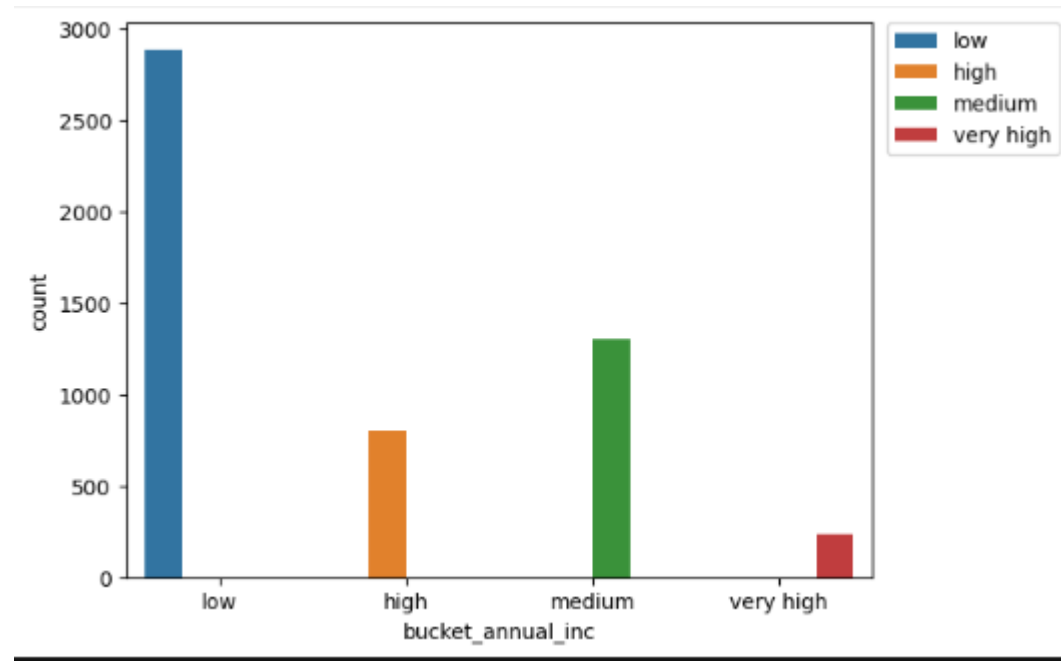
CA addr_state has high default rates

debt_conslidation purpose has high default rates irresptive of income earned

Observation 60 months tenure has high default rates. Almost 35% of the 60 month tenure loans are defaulted

Correlation :

Observations

    1) loan amount, funded amount , funded_amount_inv are deeply correlated

    2) Interest rate is negatively corelated to loan paid.

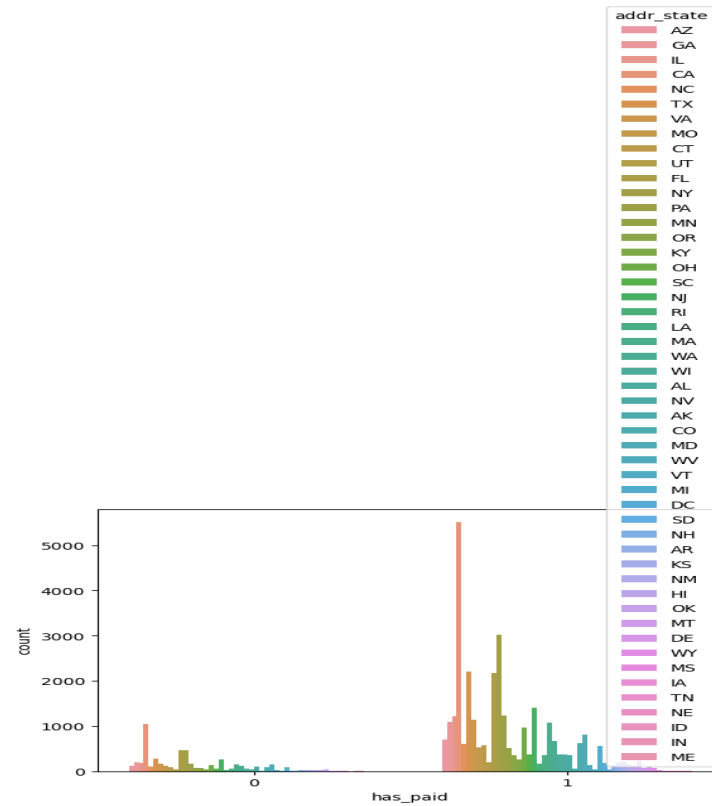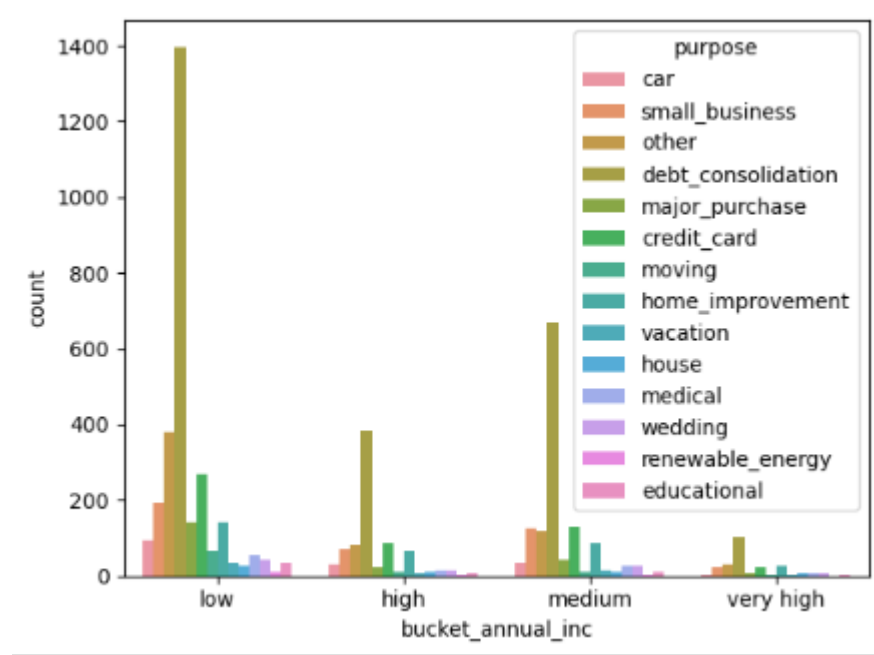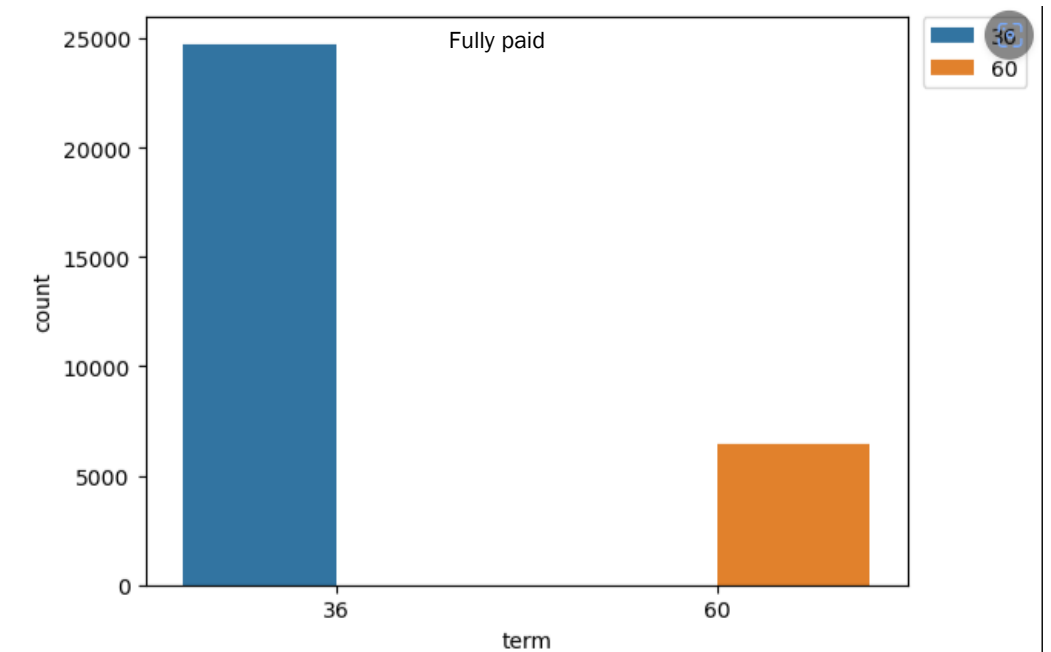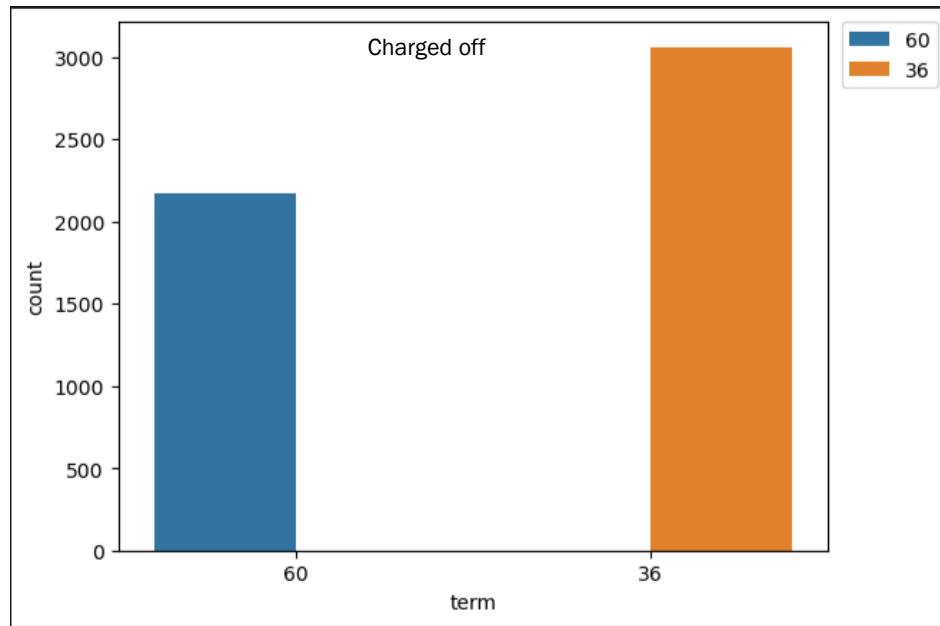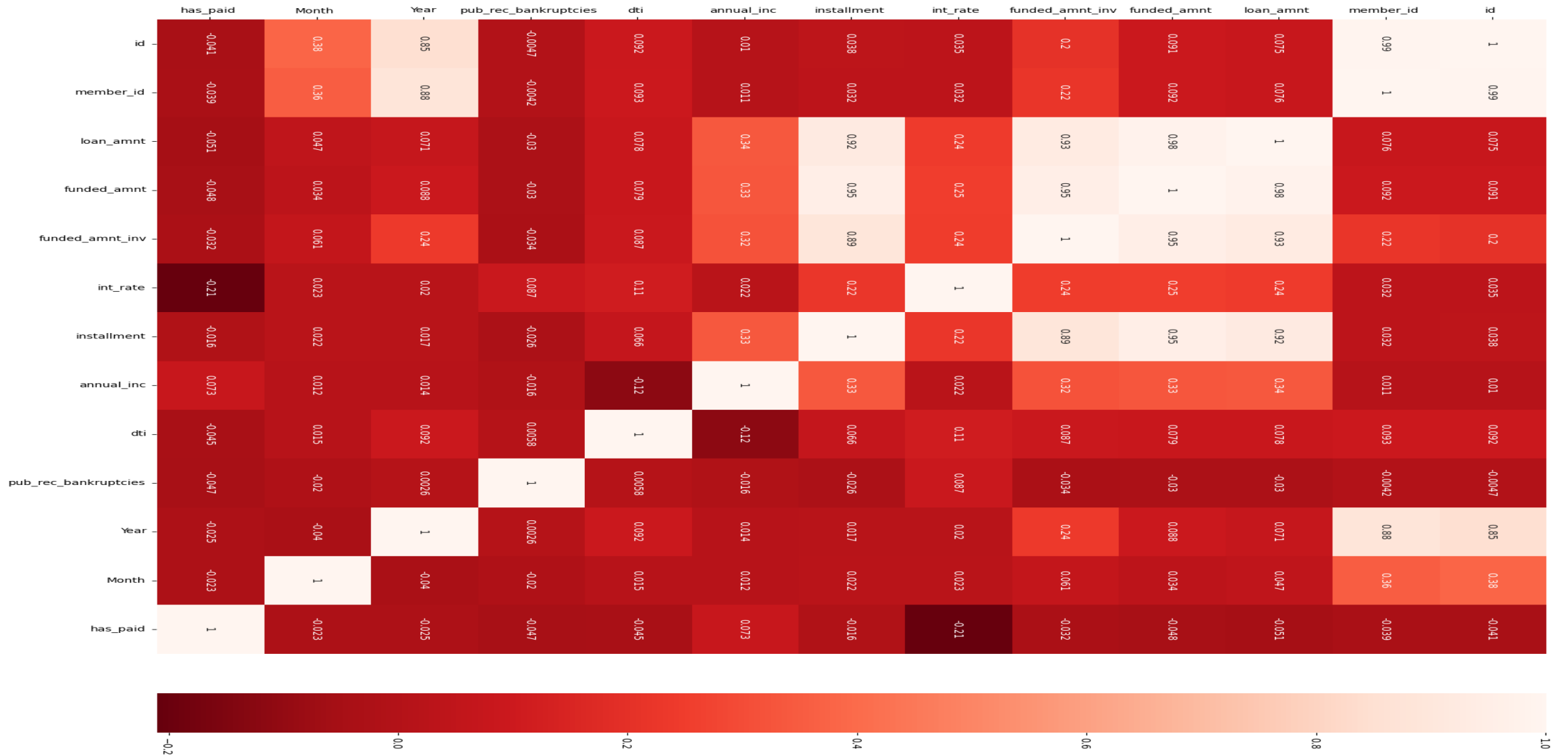|	has_paid	Month	Year	pub_rec_bankruptcies	dti	annual_inc	installment	int_rate	funded_amnt_inv	funded_amnt	loan_amnt	member_id	id
id	-0.041	0.38	0.85	-0.0047	0.092	0.01	0.038	0.035	0.2	0.091	0.075	0.99	1
member_id	-0.039	0.36	0.88	-0.0042	0.093	0.011	0.032	0.032	0.22	0.092	0.076	1	0.99
loan_amnt	-0.051	0.047	0.071	-0.03	0.078	0.34	0.92	0.24	0.93	0.98	1	0.076	0.075
funded_amnt	-0.048	0.034	0.088	-0.03	0.079	0.33	0.95	0.25	0.95	1	0.98	0.092	0.091
funded_amnt_inv	-0.032	0.061	0.24	-0.034	0.087	0.32	0.89	0.24	1	0.95	0.93	0.22	0.2
int_rate	-0.21	0.023	0.02	0.087	0.11	0.022	0.22	1	0.24	0.25	0.24	0.032	0.035
installment	-0.016	0.022	0.017	-0.026	0.066	0.33	1	0.22	0.89	0.95	0.92	0.032	0.038
annual_inc	0.073	0.012	0.014	-0.016	-0.12	1	0.33	0.022	0.32	0.33	0.34	0.011	0.01
dti	-0.045	0.015	0.092	0.0058	1	-0.12	0.066	0.11	0.087	0.079	0.078	0.093	0.092
pub_rec_bankruptcies	-0.047	-0.02	0.0026	1	0.0058	-0.016	-0.026	0.087	-0.034	-0.03	-0.03	-0.0042	-0.0047
Year	-0.025	-0.04	1	0.0026	0.092	0.014	0.017	0.02	0.24	0.088	0.071	0.88	0.85
Month	-0.023	1	-0.04	-0.02	0.015	0.012	0.022	0.023	0.061	0.034	0.047	0.36	0.38
has_paid	1	-0.023	-0.025	-0.047	-0.045	0.073	-0.016	-0.21	-0.032	-0.048	-0.051	-0.039	-0.041

| -0.2 | 0.0 | 0.2 | 0.4 | 0.6 | 0.8 | 1.0 |

Conclusion

1. Higher the Grade , higher the defaults.
2. Overall data says that most of them are repaying the loans.
3. High interest leads to more defaults.
4. CA addr_state has high default rates.
5. Observation 60 months tenure has high default rates. Almost 35% of the 60 month tenure loans are defaulted.
6. People with less experiance (less than two years) Very high experience (10 + years) are prone to high defaults.
7. People who are in Rents and Mortgauge Home_Ownership are prone to high defaults.
8. More defaults are oversved in December months. Defaults are constantly increasing over the years.
9. Low income earners are likely to default.
10. debt_conslidation purpose has high default rates irresptive of income earned.
11. loan amount, funded amount , funded_amount_inv are deeply correlated.
12. Interest rate is negitively corelated to loan paid.

13. Higher loan_amount are likely to be charged off. Low and Medium loan amounts are likely to be fully paid off

Thank you ☺