Lending Club Case Study

Exploratory Data Analysis

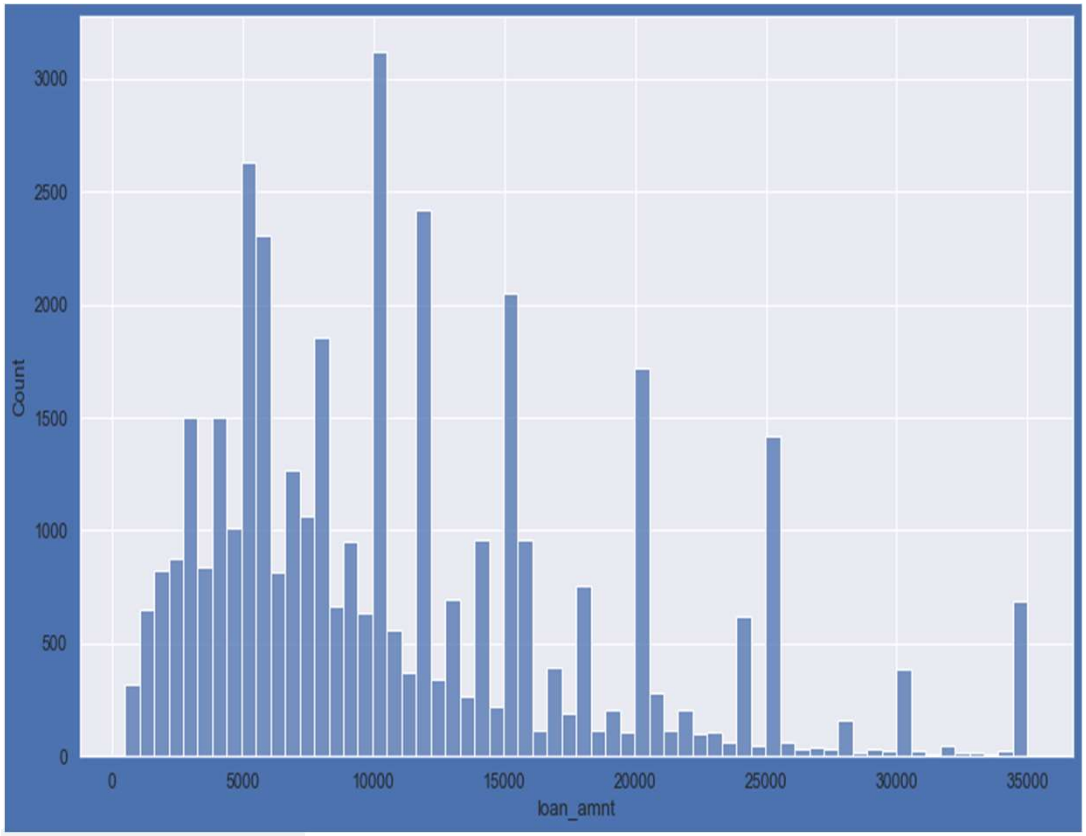
Contents

Overview & Problem Statement

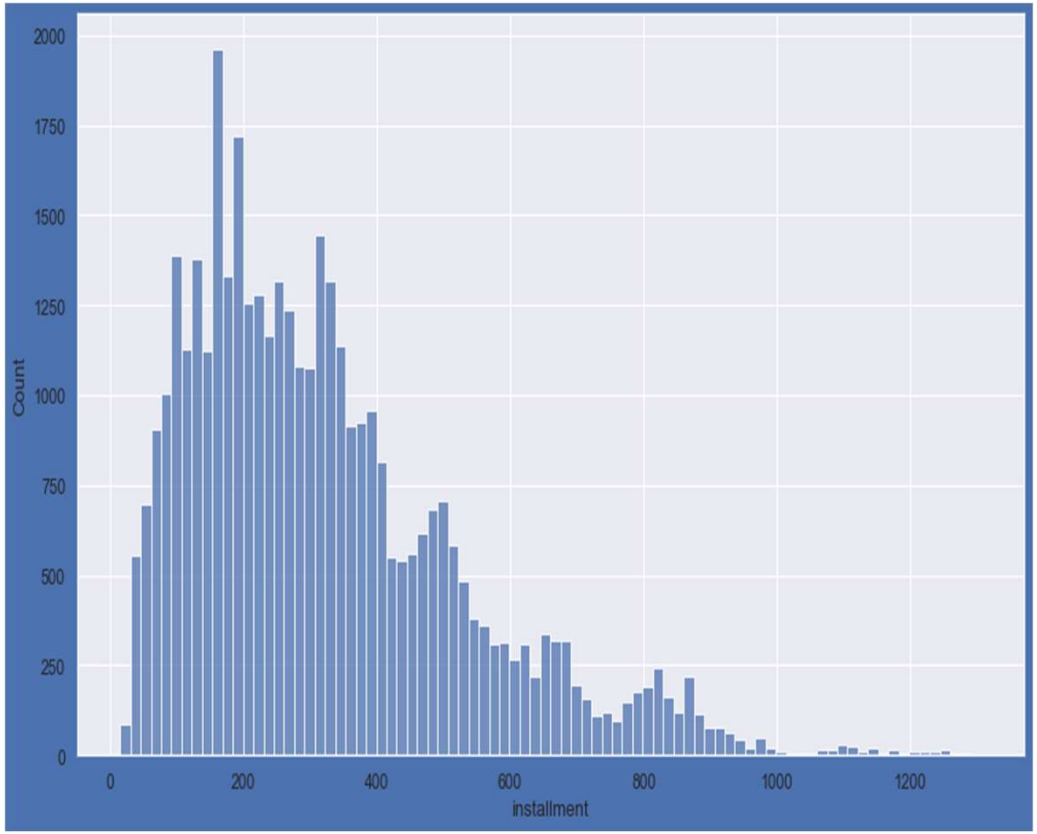
- **A consumer finance company specializes in various types of loans to urban customers**
- **When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile**
- **Two types of risk are associated with the bank's decision:**
  - If the 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
- **The objective of the analysis is to use the information about past loan applicants and find whether they defaulted or not by identifying patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.**

Univariate Analysis – Numeric Variables

The distribution of loan amount is mostly between 2500 and 7500

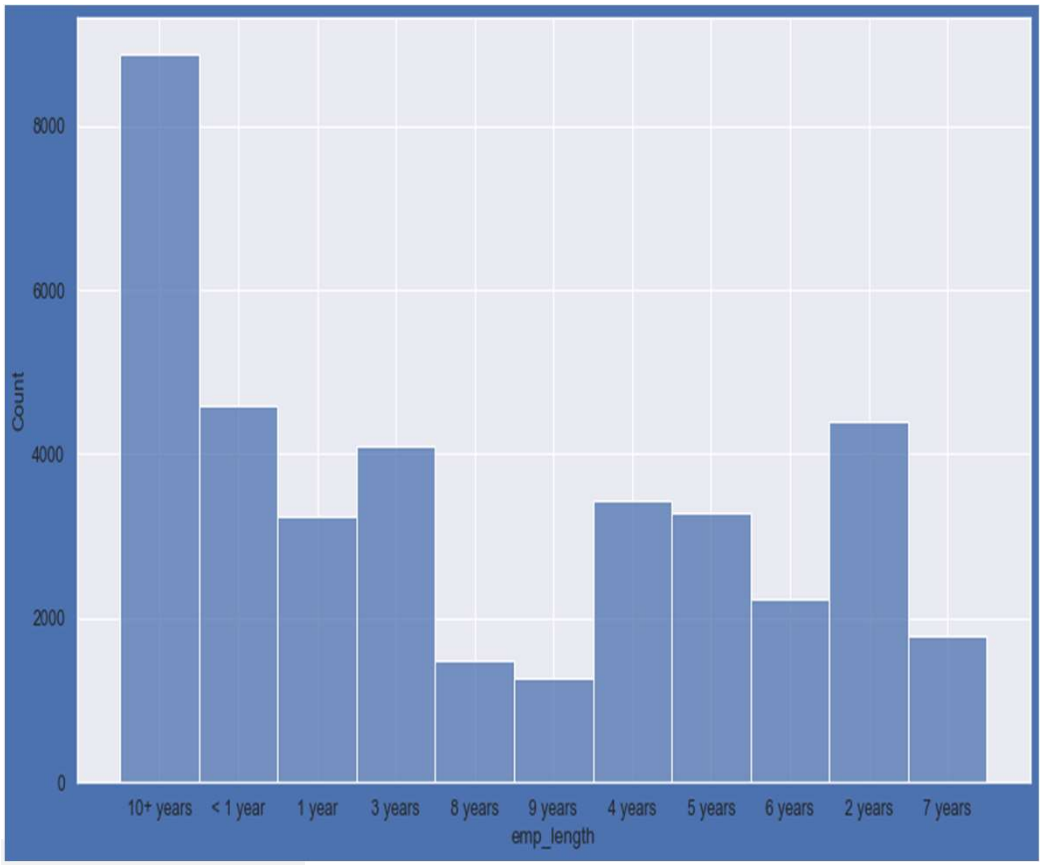


The distribution of instalments is mostly between 100 to 300

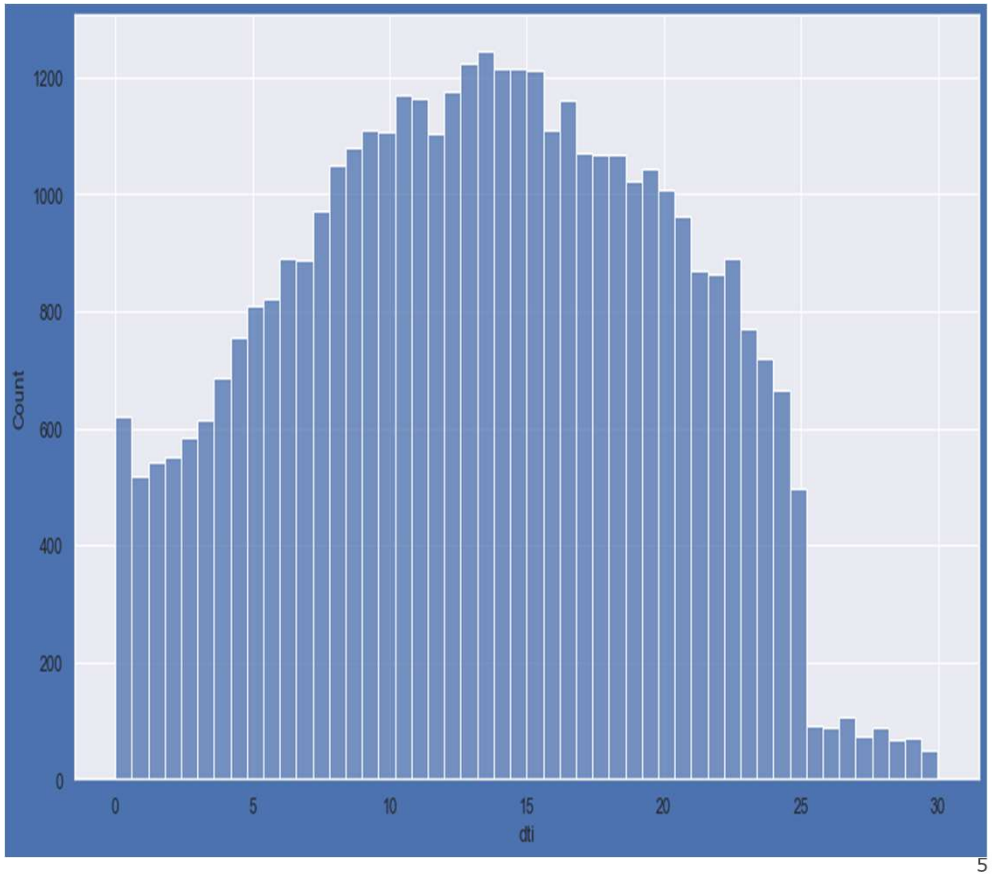


Univariate Analysis – Numeric Variables

Employment length is mostly above 10+ years

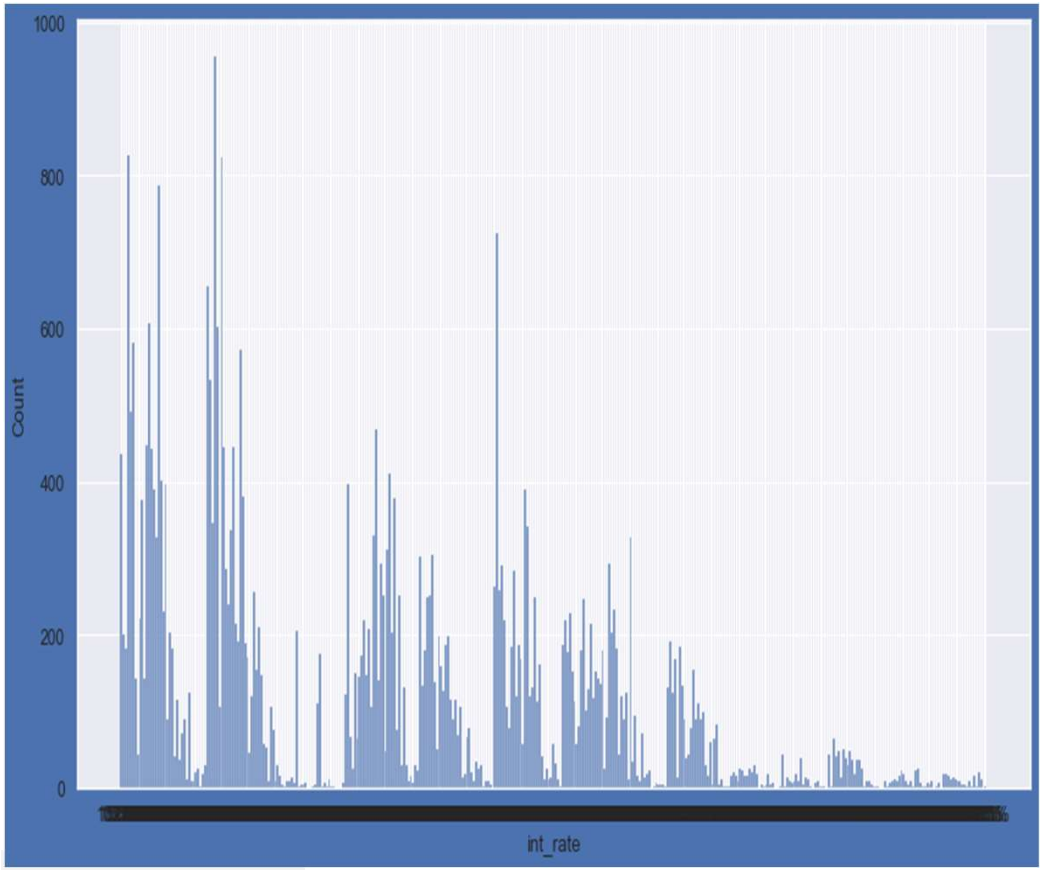


The distribution of dti is between 7.5 to 17.5

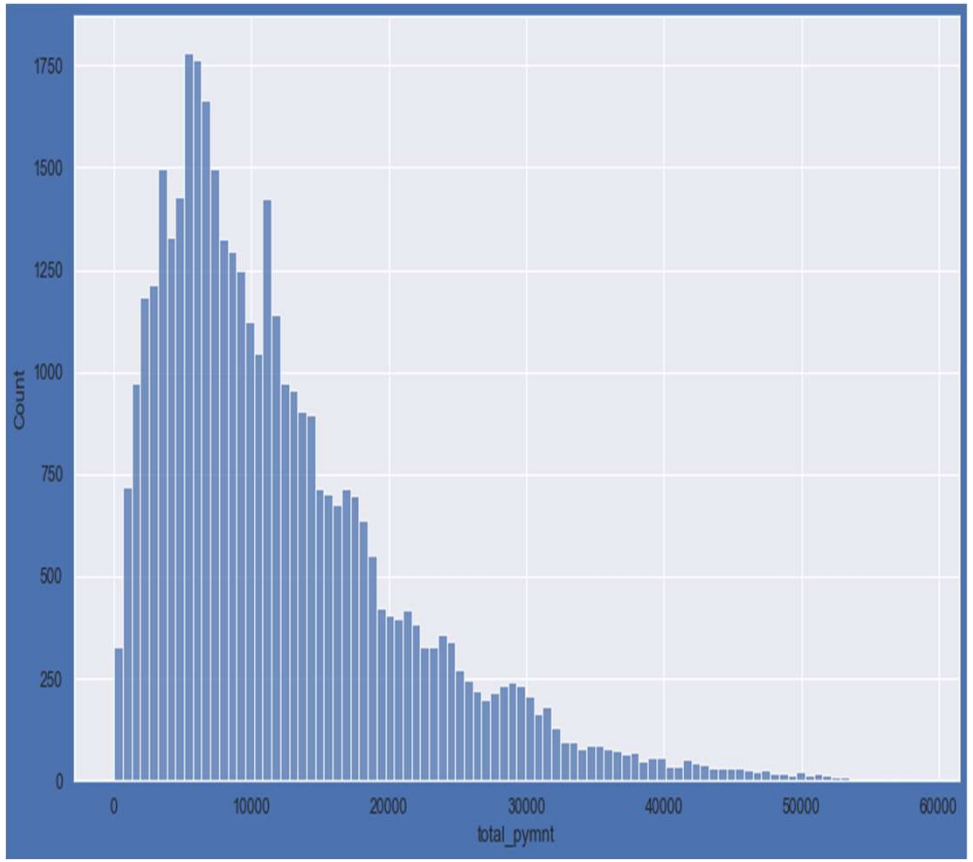


Univariate Analysis – Numeric Variables

Most of the interest rates are between 10 to 15%

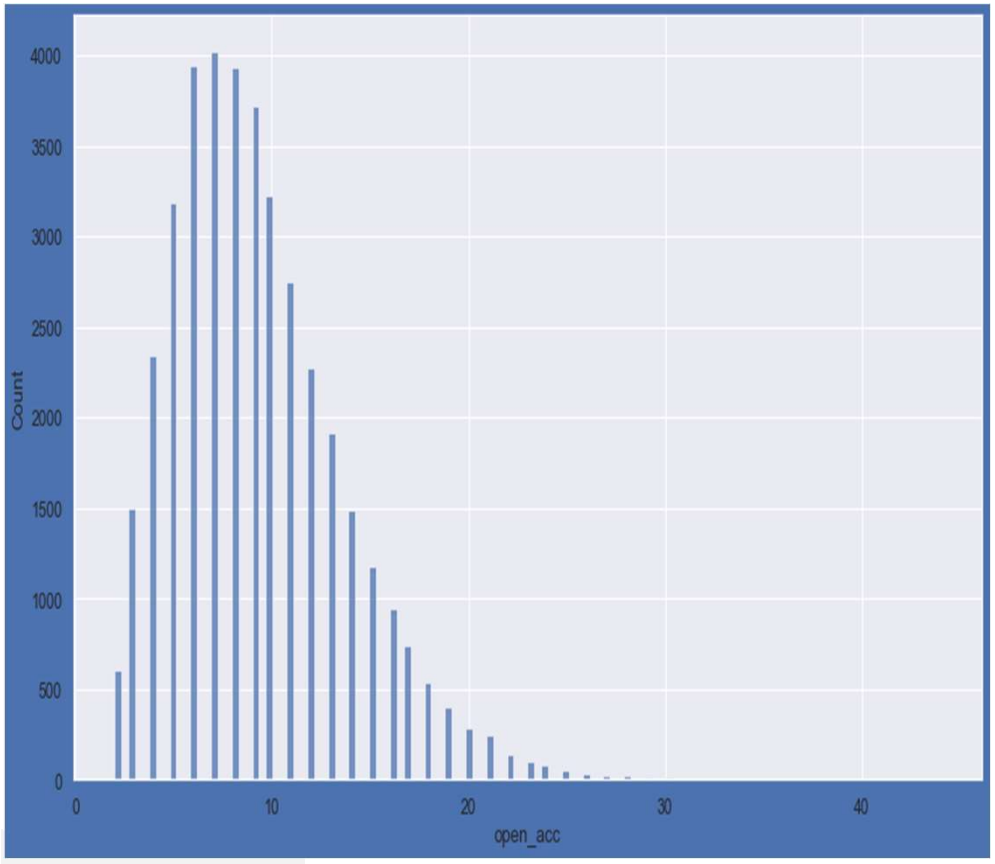


Total Payment and Total Payment Investors is between 4K to 12K



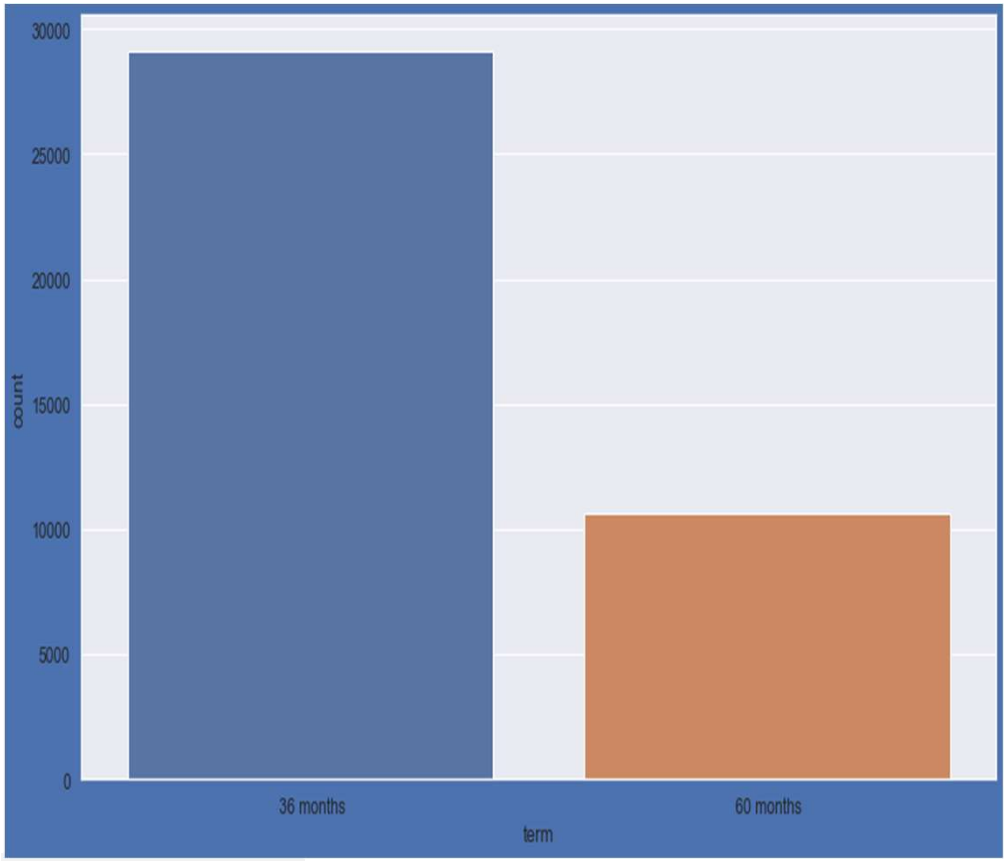
Univariate Analysis – Numeric Variables

The number of open credit lines in the borrower's credit file are mostly between 4 to 12

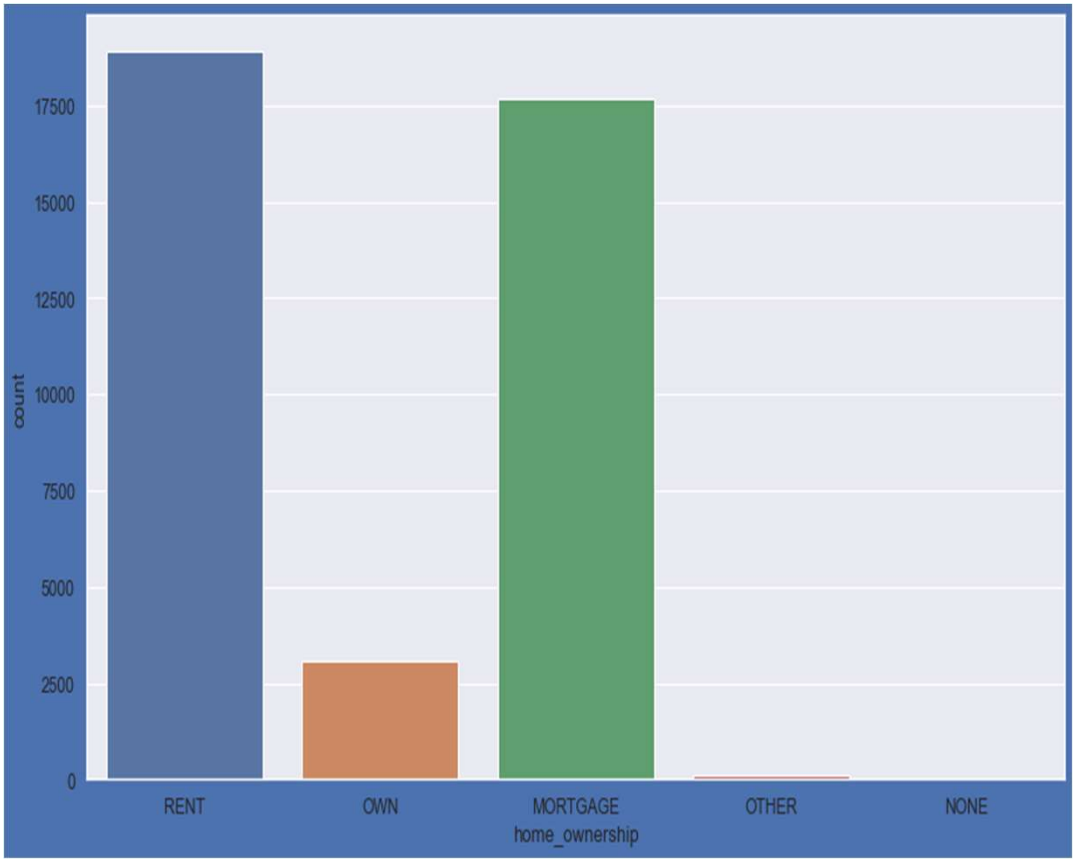


Univariate Analysis – Categorical Variables

Most of the customers are having a term of 36 months

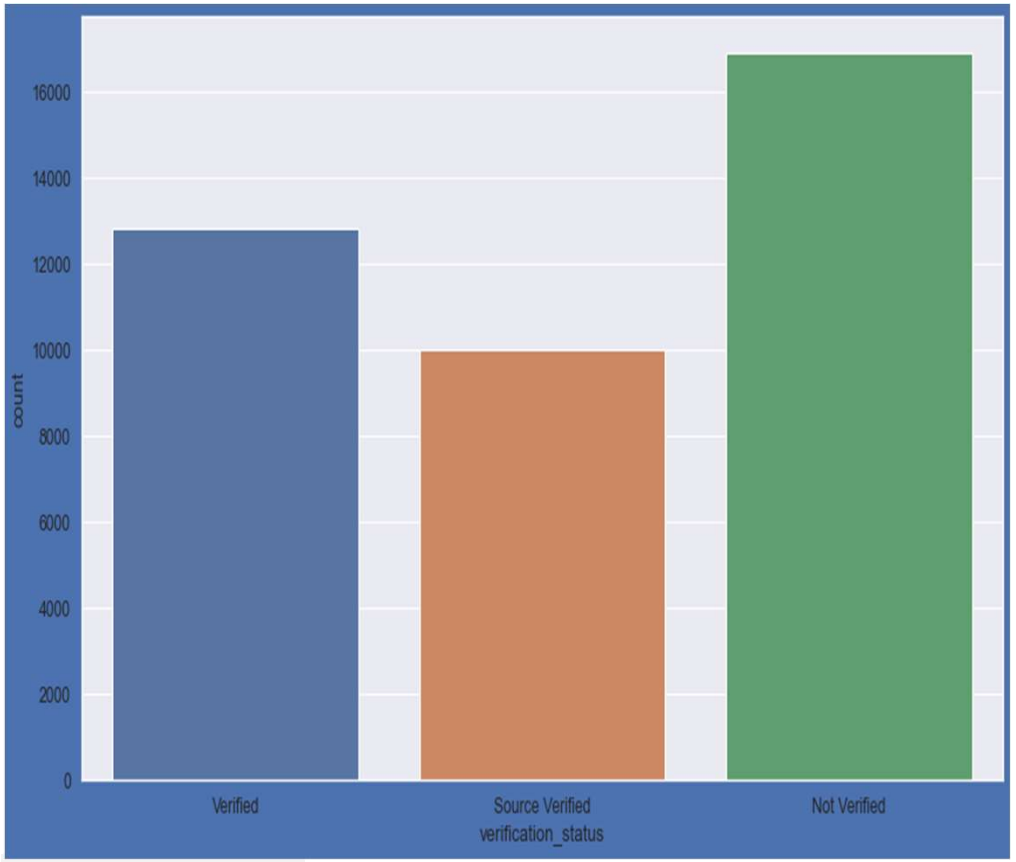


Most of the customers are living on rent or mortgage

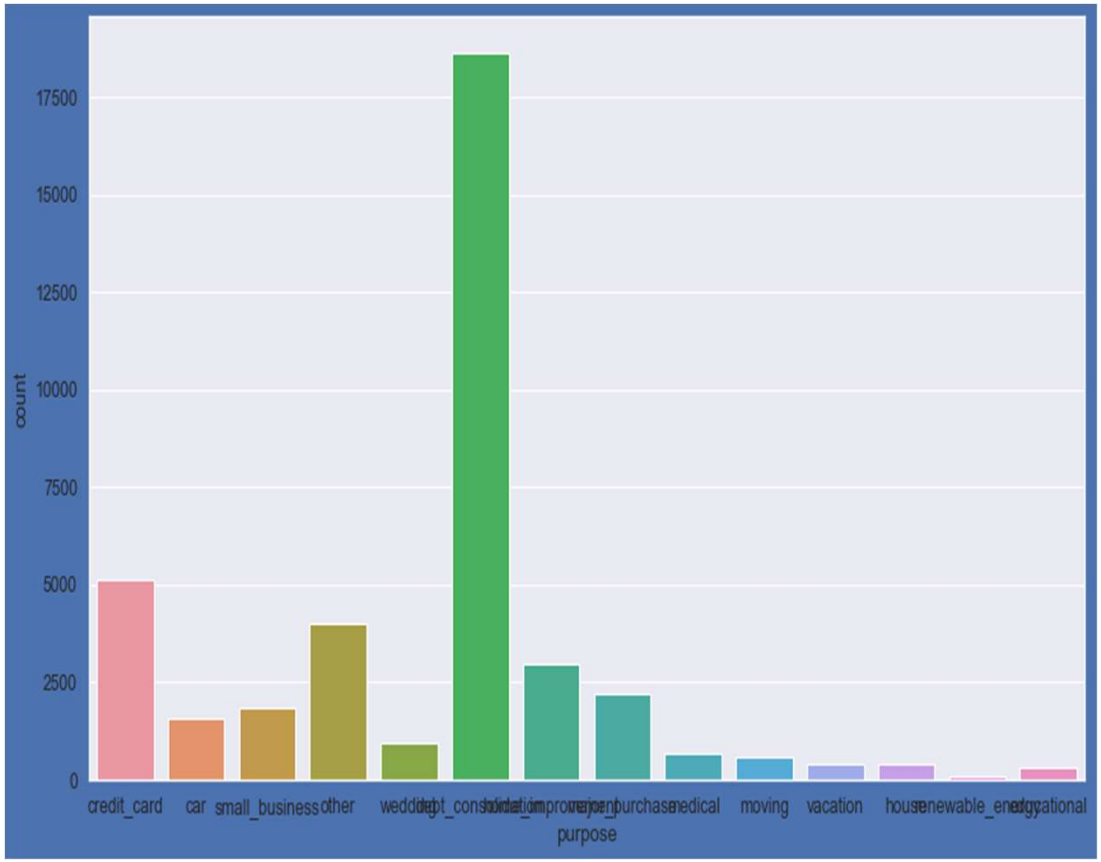


Univariate Analysis – Categorical Variables

Most of the customers are in Not Verifed status



Most of the customers have taken the loan for debt consolidation & credit bill



Bivariate & Segmented Analysis

1) Annual Income to Debt To Income Ratio i.e dti are negatively correlated

2) Loan Amount,Investor Amount and Funding Amount are strongly correlated

3) Positive correlation between Annual Income and employment years

4) Positive correlation between annual income and funded amount that means people with high income gets high funded amount

5) Positive correlation between annual income and total payment

	id	loan_amnt	funded_amnt	funded_amnt_inv	int_rate	emp_length	annual_inc	dti	delinq_2yrs	inq_last_6mths
id	1.000000	0.113729	0.124942	0.228398	0.049439	0.094223	0.035617	0.096267	-0.007051	-0.043228
loan_amnt	0.113729	1.000000	0.982261	0.936088	0.283147	0.157574	0.393921	0.102003	-0.040590	0.002596
funded_amnt	0.124942	0.982261	1.000000	0.953811	0.289135	0.156984	0.390923	0.101006	-0.040352	0.003009
funded_amnt_inv	0.228398	0.936088	0.953811	1.000000	0.282716	0.165855	0.373004	0.108636	-0.045112	-0.012028
int_rate	0.049439	0.283147	0.289135	0.282716	1.000000	0.008397	0.046231	0.106781	0.151657	0.134051
emp_length	0.094223	0.157574	0.156984	0.165855	0.008397	1.000000	0.224709	0.060325	0.013775	0.011293
annual_inc	0.035617	0.393921	0.390923	0.373004	0.046231	0.224709	1.000000	-0.048269	0.023511	0.030806
dti	0.096267	0.102003	0.101006	0.108636	0.106781	0.060325	-0.048269	1.000000	-0.030297	0.008442
delinq_2yrs	-0.007051	-0.040590	-0.040352	-0.045112	0.151657	0.013775	0.023511	-0.030297	1.000000	0.009636
inq_last_6mths	-0.043228	0.002596	0.003009	-0.012028	0.134051	0.011293	0.030806	0.008442	0.009636	1.000000
open_acc	0.017315	0.162853	0.160771	0.149076	-0.016446	0.097067	0.266462	0.301169	0.015069	0.092330
pub_rec	-0.016408	-0.042522	-0.042774	-0.045010	0.105480	0.056844	-0.000341	-0.008468	0.009243	0.025566
revol_bal	0.008012	0.314861	0.306918	0.287817	0.071023	0.160434	0.373919	0.284960	-0.068621	-0.024325
total_acc	0.038047	0.234502	0.228892	0.222555	-0.073242	0.204958	0.382717	0.252085	0.066839	0.107837

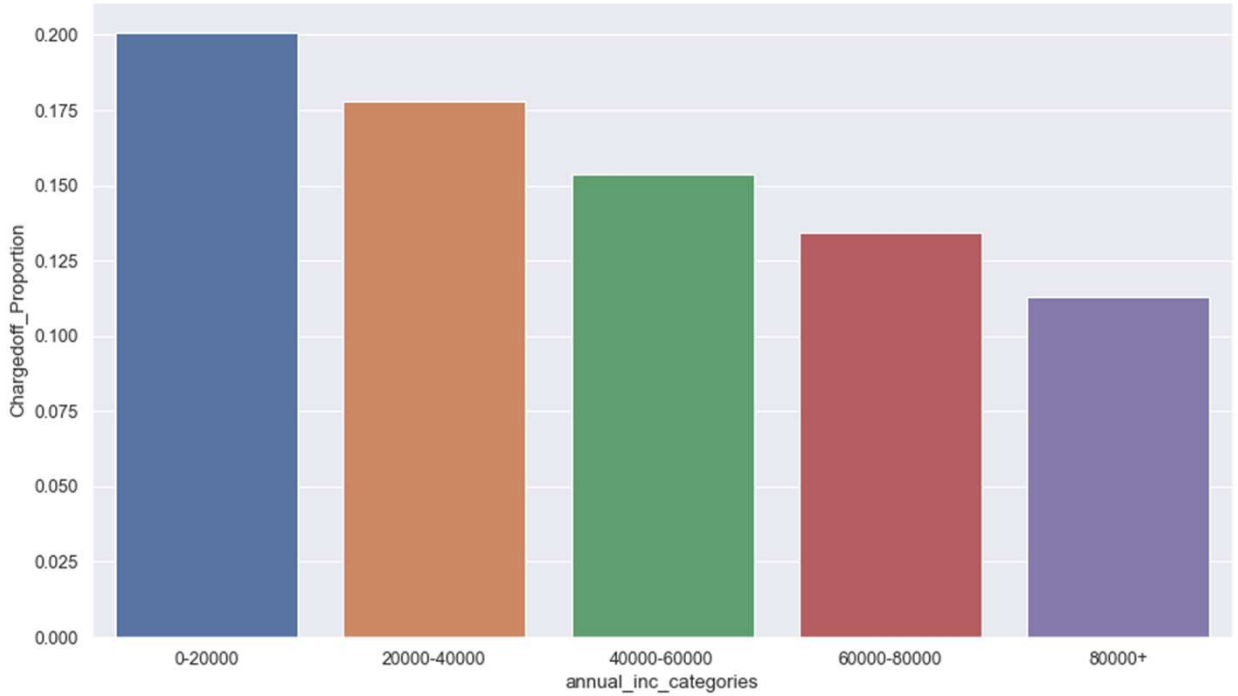
Bivariate & Segmented Analysis

Insights on Annual Income vs Loan Status

1) Income range 80000+ has less chances of charged off.

2) Income range 0-20000 has high chances of charged off.

3) With increase in annual income charged off proportion gets decreased. So, they are inversely proportional.

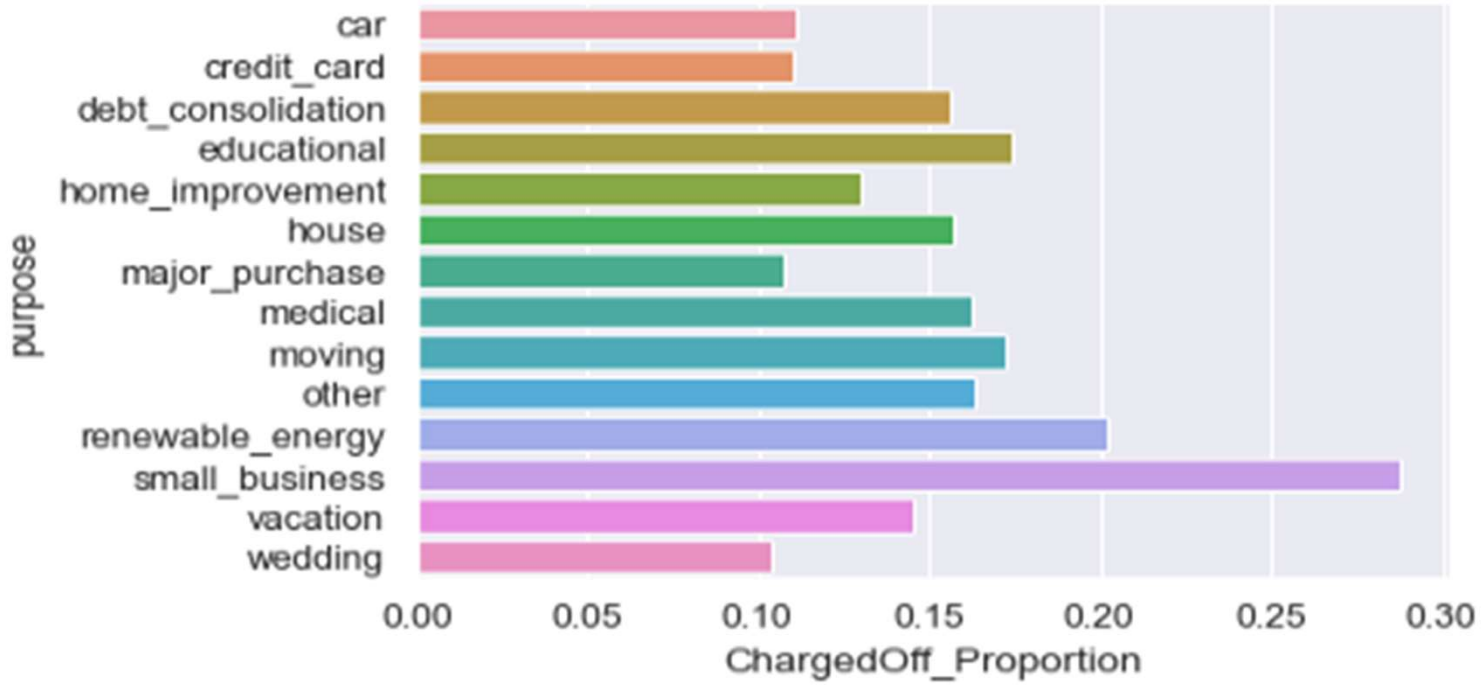


Bivariate & Segmented Analysis

Insights on Purpose of Loan vs Charged Off Proportion

1) Small Business applicants have high chances of getting charged off.

2) Renewable_energy charged off proportion is better as compared to other categories



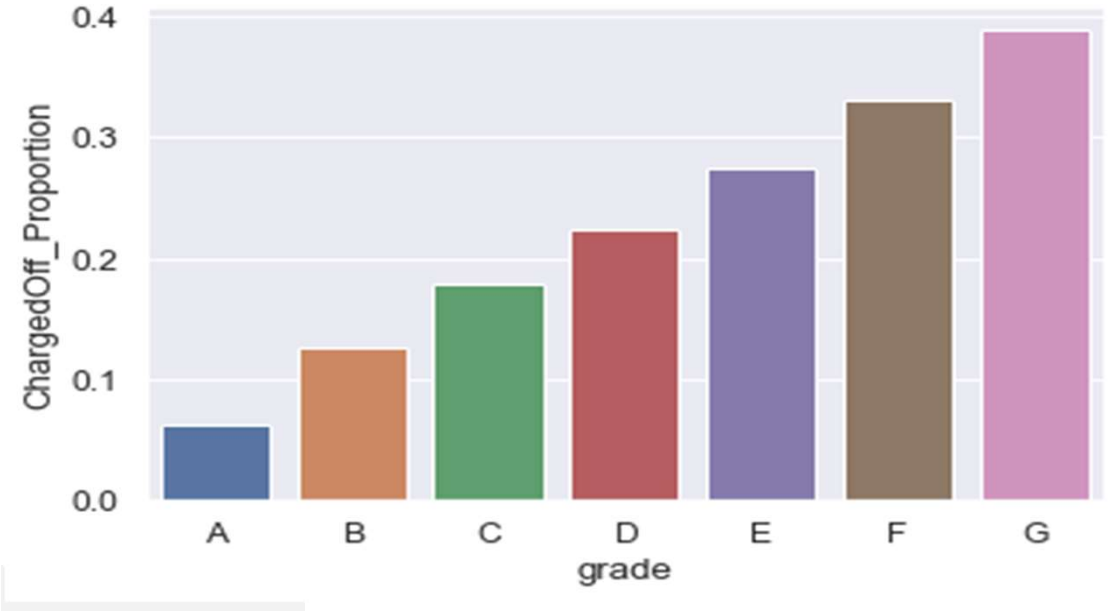
Bivariate & Segmented Analysis

Insights on Grade vs Charged Off Proportion

1) Grade A has least chances of getting charged off

2) Grade F and Grade G have high chances of getting charged off

3)Chances of getting charged off is increasing with grades moving from A to G

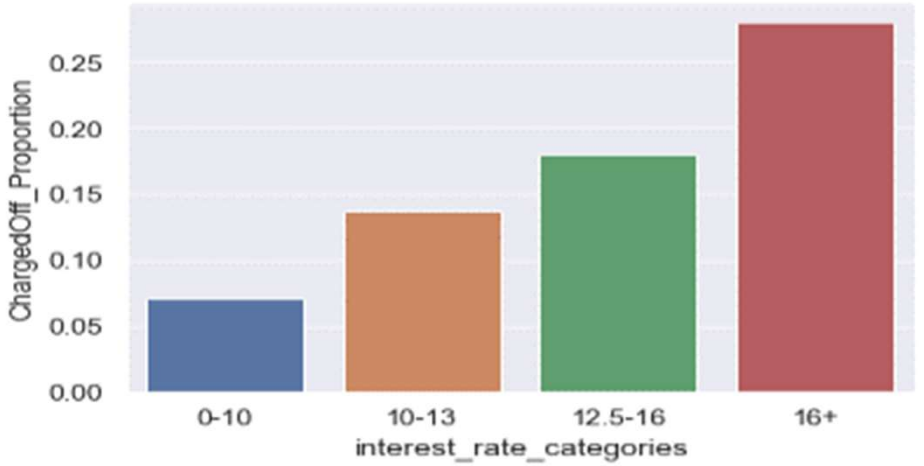


Insights on Interest Rate vs Charged Off Proportion

1.Interest Rates which are less than 10% have very less chances of charged off

2.Interest Rates greater than 16 have higher chances of charged off

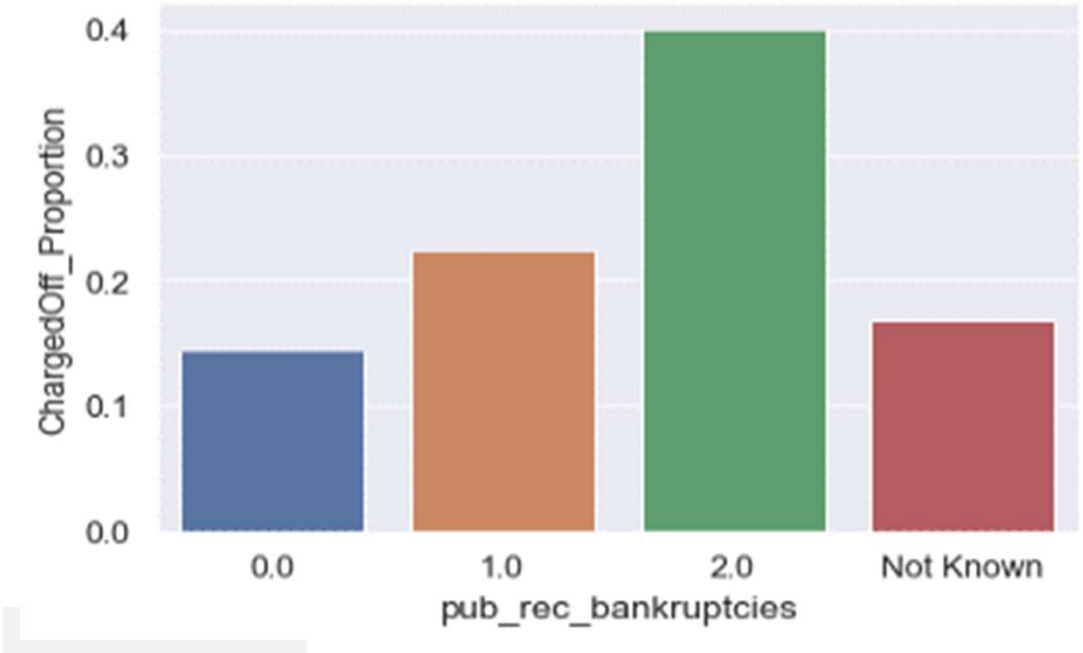
3.Charged Off Proportion increases with higher interest rates



Bivariate & Segmented Analysis

Insights on pub_rec_bankruptcies vs charged off proportion

#1) Those who already have pub_rec_bankruptcies value 1, have charged off proportion higher than those who have no pub_rec_bankruptcies



Insights on Verification Status vs Charged Off Proportion

1) Not Verified Status have higher chances of getting charged off when compared to verified status

Bivariate & Segmented Analysis

Insights on Address State vs Charged Off Proportion

1)State NE has very high chances of charged off but number of applications are too low to make any decisions

2) States NV,CA and FL states shows good number of charged offs with good number of applications

]:

loan_status	addr_state	Charged Off	Fully Paid	Total	ChargedOff_Proportion
28	NE	3.0	2.0	5.0	0.600000
32	NV	98.0	337.0	435.0	0.225287
13	ID	1.0	4.0	5.0	0.200000
0	AK	13.0	54.0	67.0	0.194030
40	SD	12.0	50.0	62.0	0.193548
9	FL	469.0	2093.0	2562.0	0.183060
11	HI	27.0	126.0	153.0	0.176471
31	NM	30.0	143.0	173.0	0.173410
24	MO	110.0	528.0	638.0	0.172414
4	CA	1022.0	5103.0	6125.0	0.166857
30	NJ	258.0	1302.0	1560.0	0.165385
10	GA	203.0	1027.0	1230.0	0.165041
36	OR	65.0	342.0	407.0	0.159705
46	WA	119.0	629.0	748.0	0.159091
20	MD	141.0	750.0	891.0	0.158249
27	NC	109.0	583.0	692.0	0.157514
43	UT	36.0	197.0	233.0	0.154506
3	AZ	116.0	656.0	772.0	0.150259

Recommendations

1)Grades is a good criteria for detecting defaulters. Company should take caution before giving loans to low grade customers

2) Small business loans default more. Hence, company should stop issuing loans to these businesses

3) Company should reduce the high interest loans for 60 months tenure as they are more prone to defaults

4) Company should stop giving loans to home ownership and mortgage as they are taking higher loans and defaulting more

5) Company should caution against giving loans to borrowers who are from CA, FL & NY to make profits

6) People with more number of public derogatory records have more chances for going bankrupt. Hence, company should check the public derogatory records for borrowers