LENDING CLUB CASE STUDY

L JENIL CHRISTO

Contents

- Problem Statement
- Approach
- Data Preparation
- Univariate Analysis
- Bivariate Analysis
- Multi-Variate Analysis
- Conclusion

Problem Statement

- Lending club a finance company which specialises 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

- The Objective of this study is to find the driving factors behind loan default, for the company to assess its risk and portfolio management

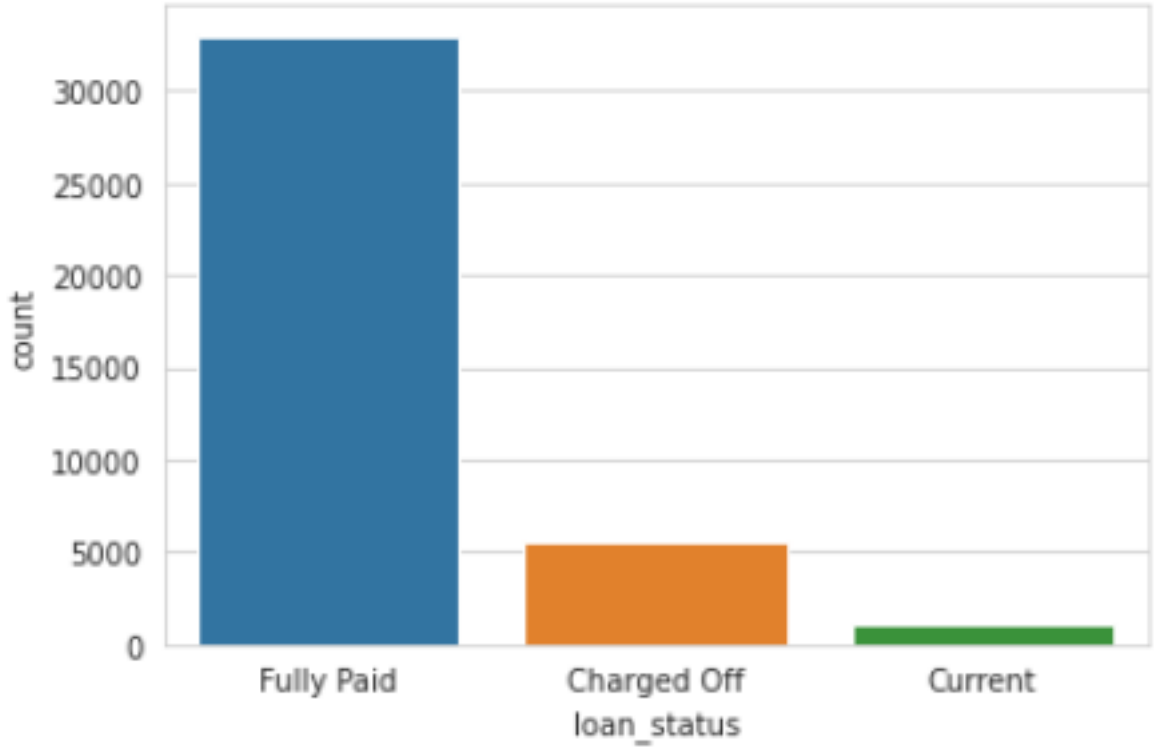
Approach

- Understand the data
- Remove missing columns
- Handle missing values
- Outlier Treatment
- Study the distribution of numeric features
- Make count plots on the charged off and fully paid loans
- Do Bivariate comparison against the target ‘loan_status’
- Do multivariate analysis on important variables

Univariate Analysis

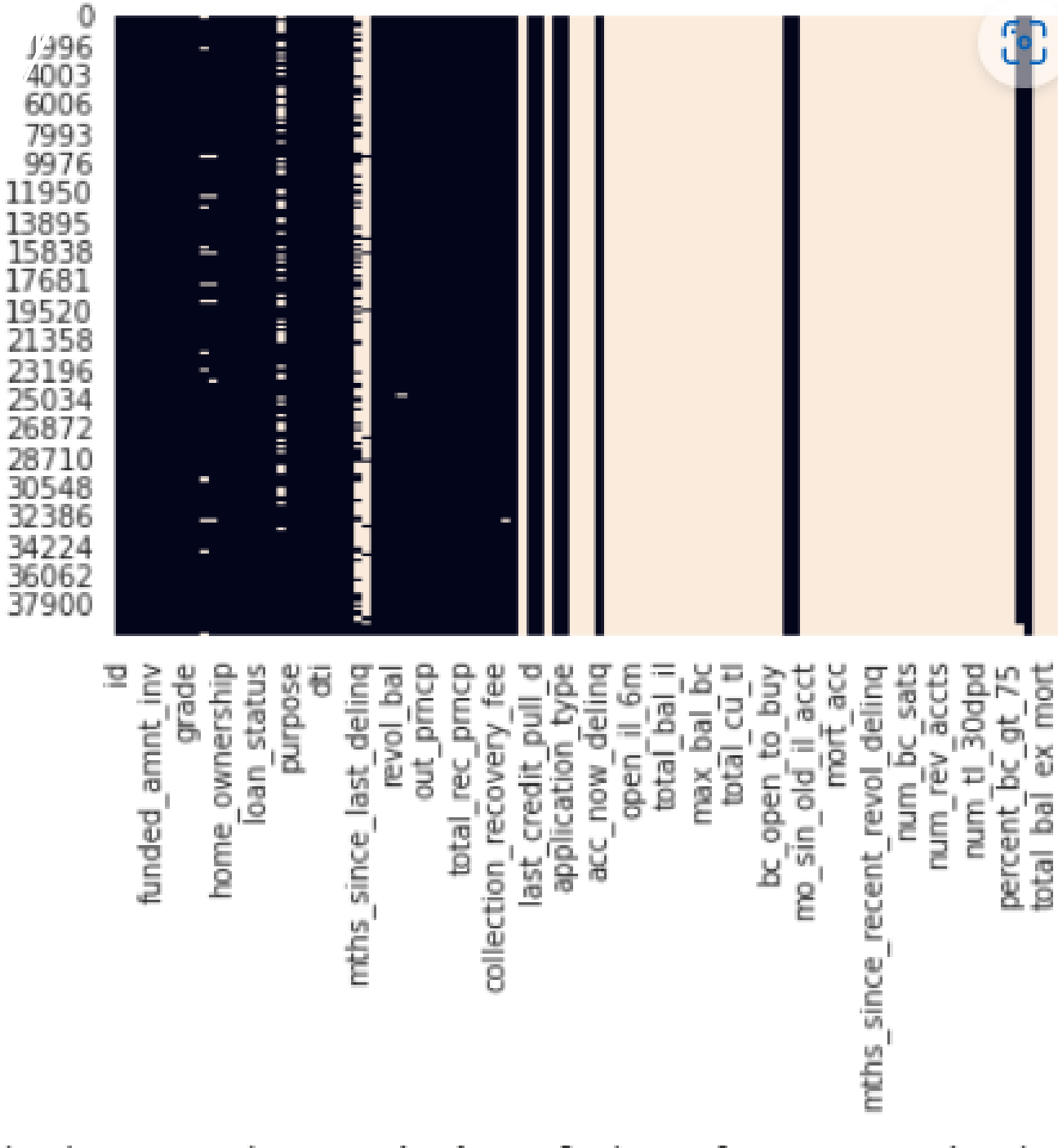
- **<u>Distribution of Target variable:</u>**
- Current loans doesn't contribute to analysis, hence
- Can be removed.

```
82.96 % of the loans are Fully Paid
14.17 % of the loans are Charged Off
2.87 % of the loans are Current loas
```



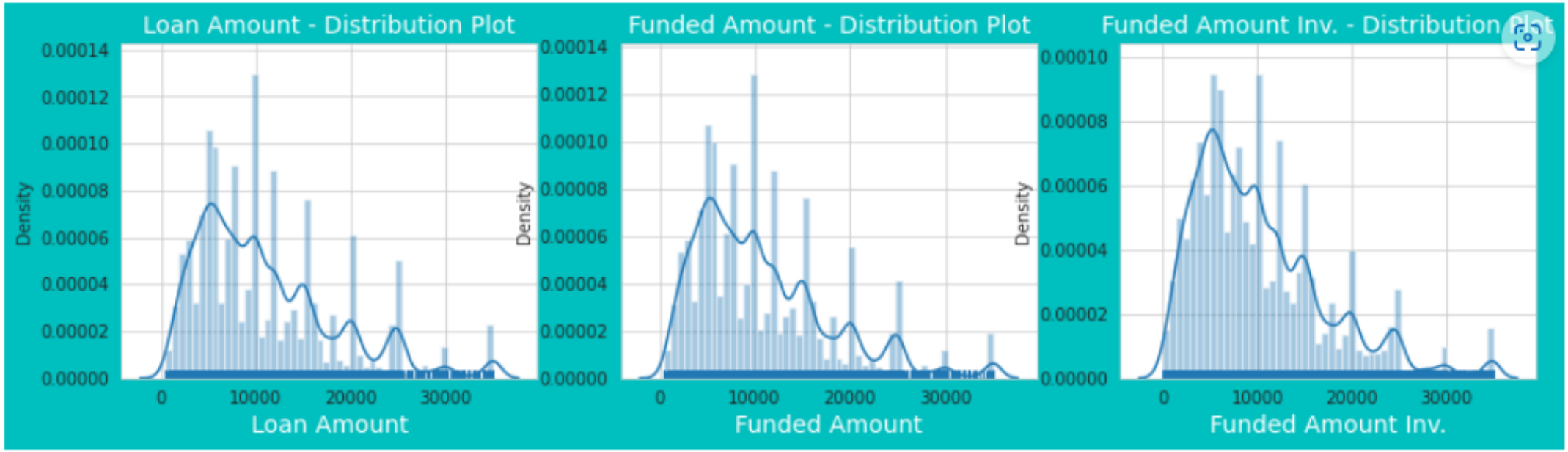
Missing values

- Remove missing values in columns. Any column with missing value less than 30% and significant to analysis, we will do imputation

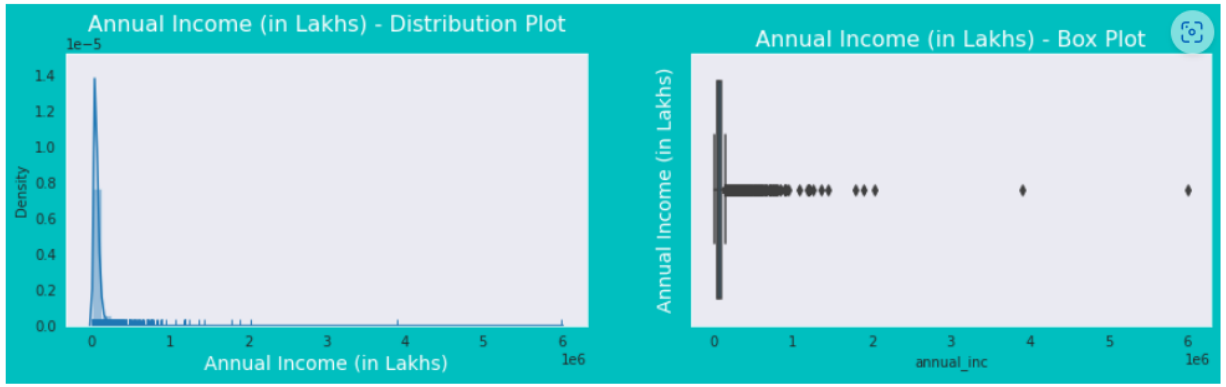


Feature Correlation

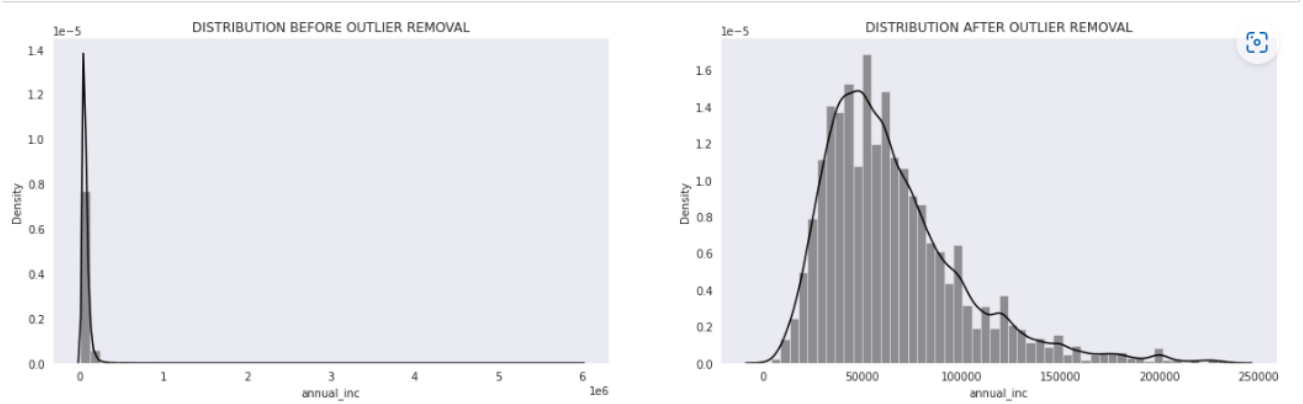
- Loan amount, Funded amount and funded amount inv have same distribution and have the same
- Correlation, hence we will use loan amount

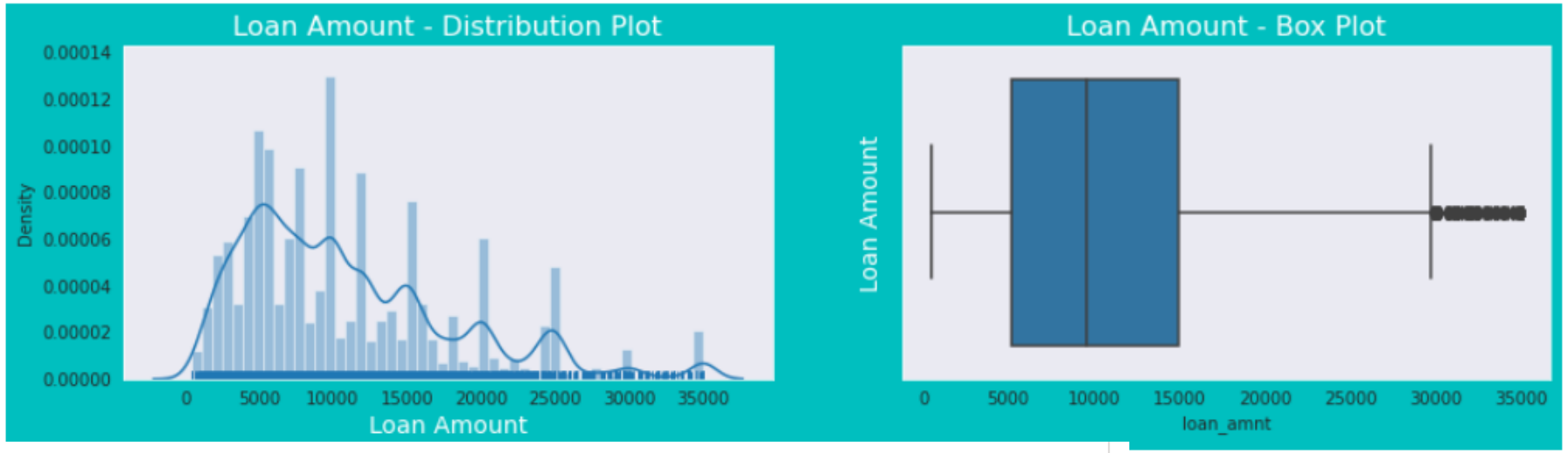


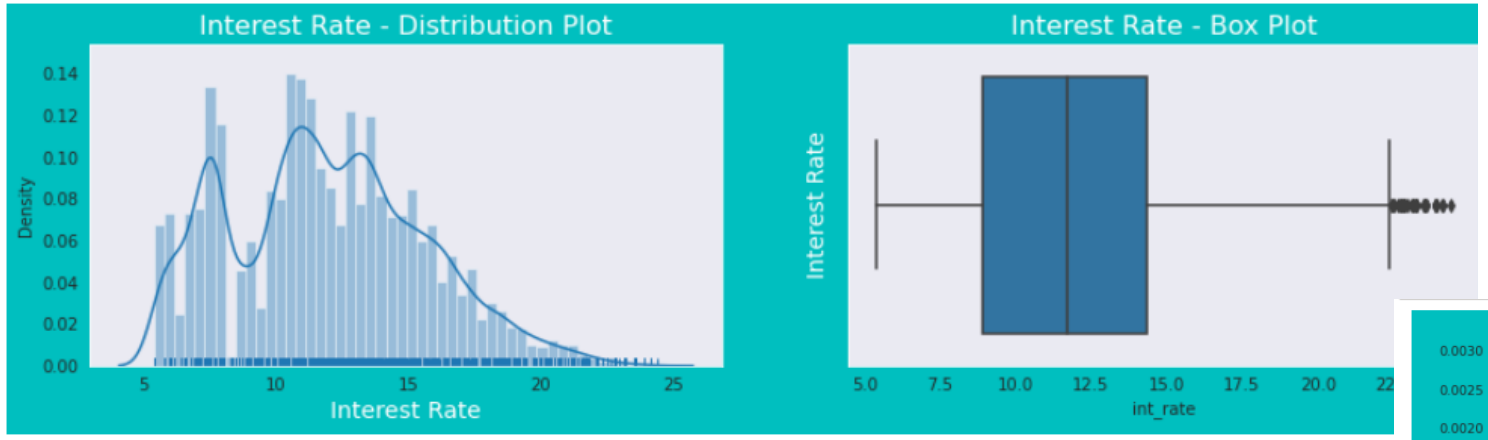
Outlier removal

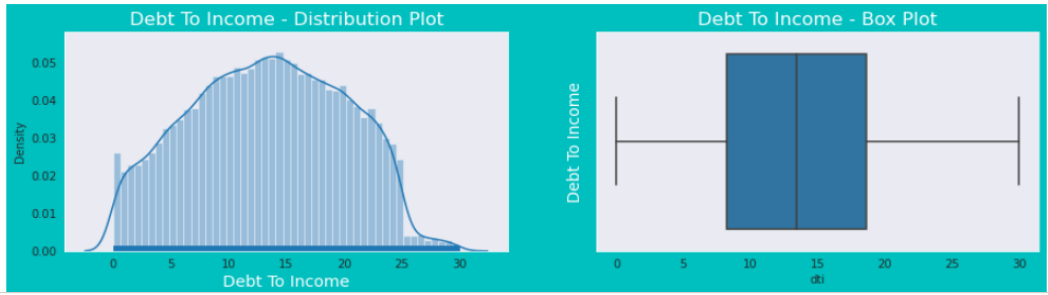


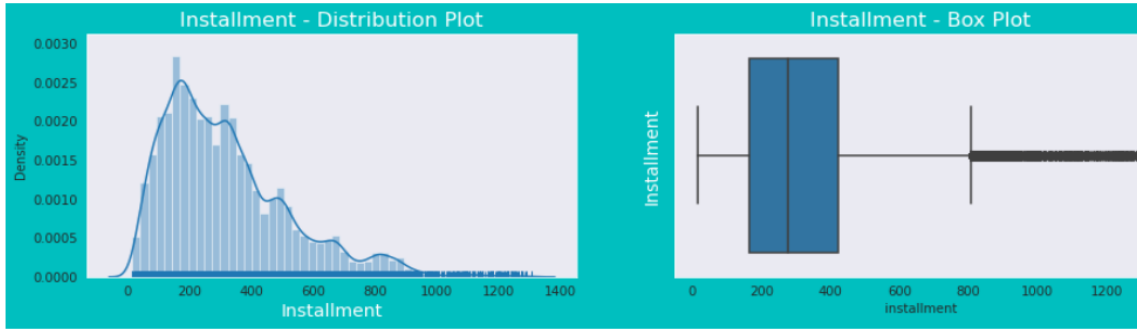
- Annual income tend to have outliers far from the upper fence. This is removed by filtering out values > 99th percentile



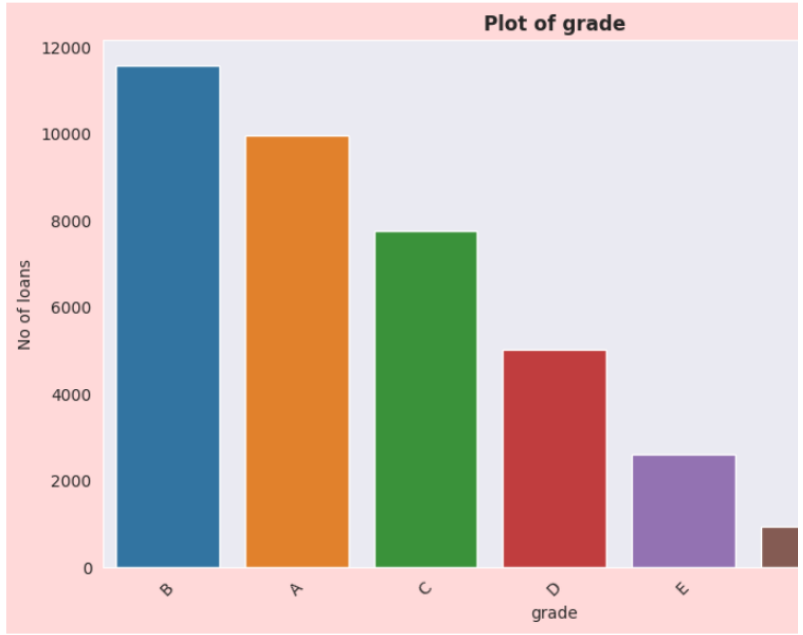


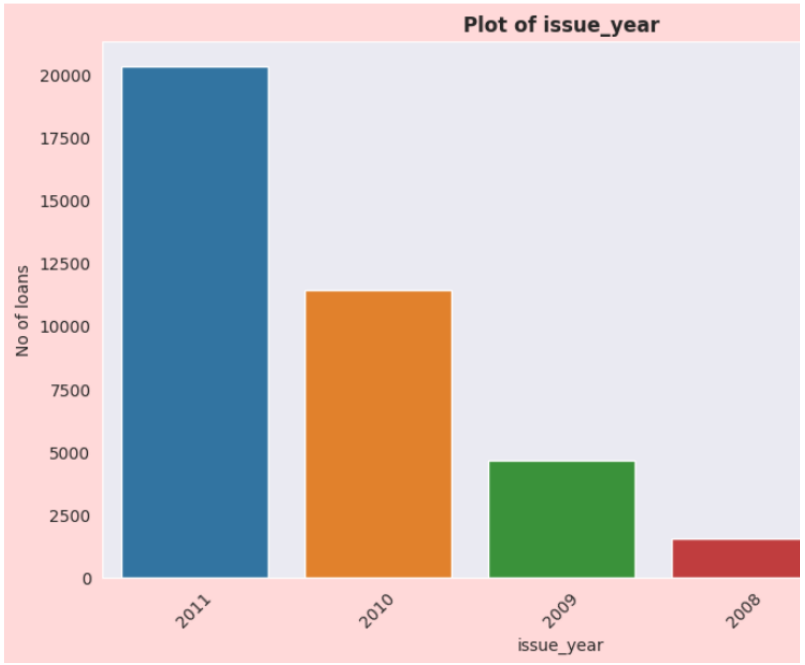


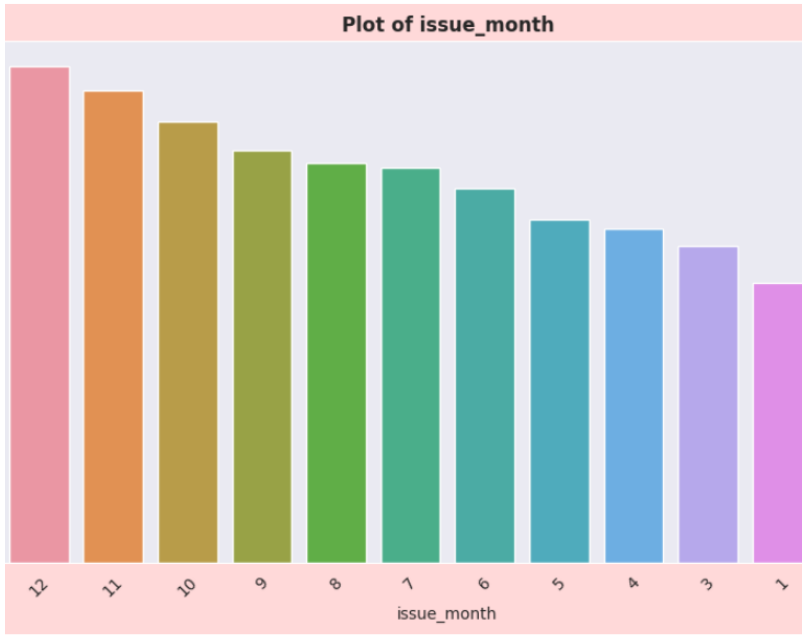


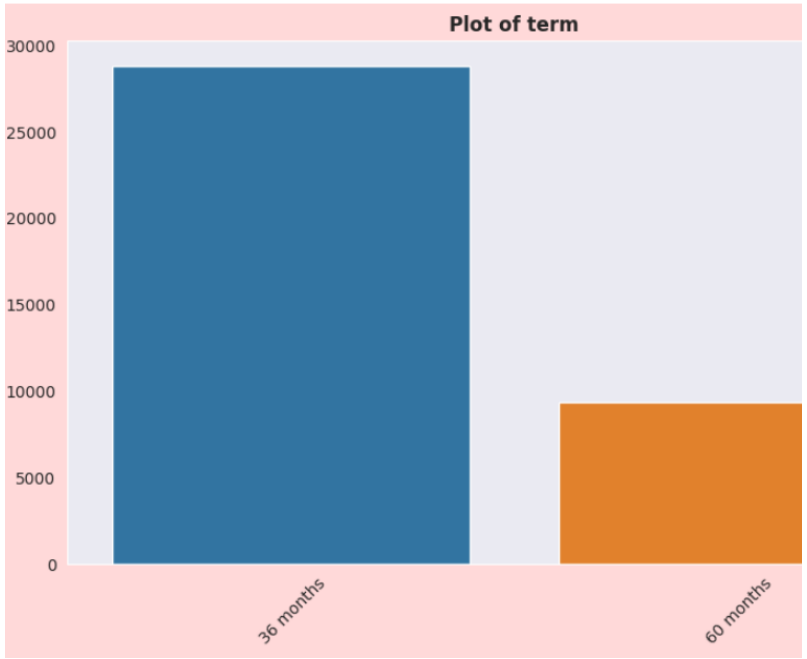


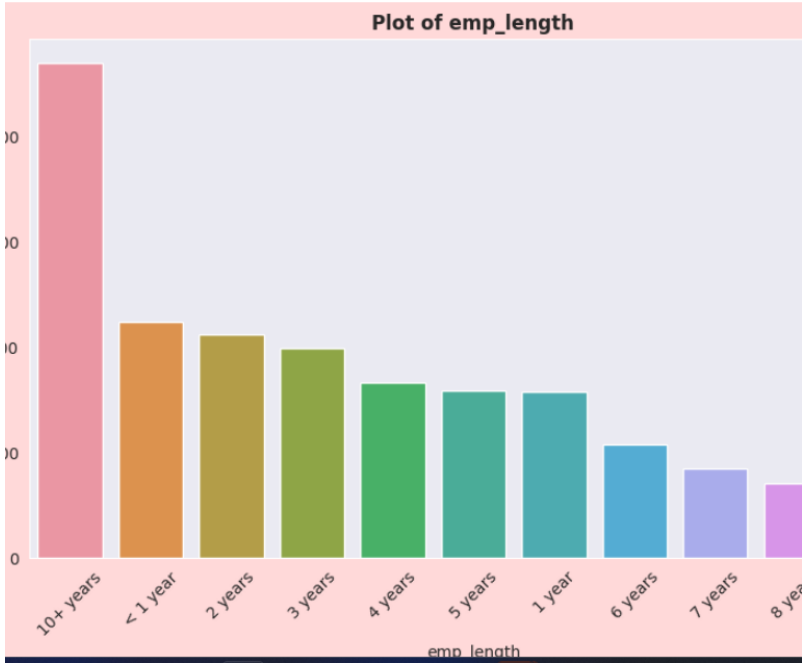
Univariate analysis


Plot of grade
No of loans
grade
B
A
C
D
E

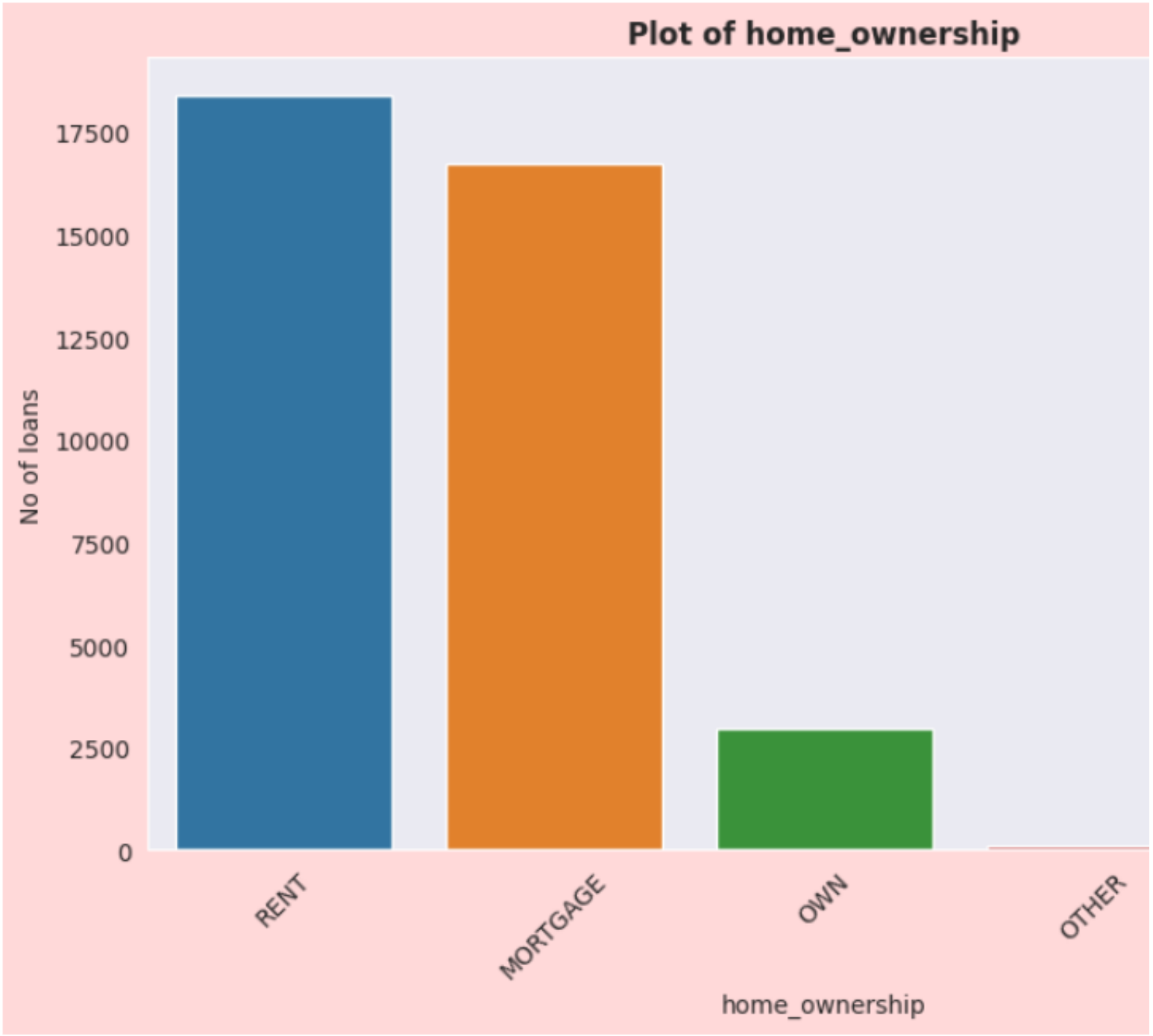

Plot of issue_year
No of loans
issue_year
2011
2010
2009
2008

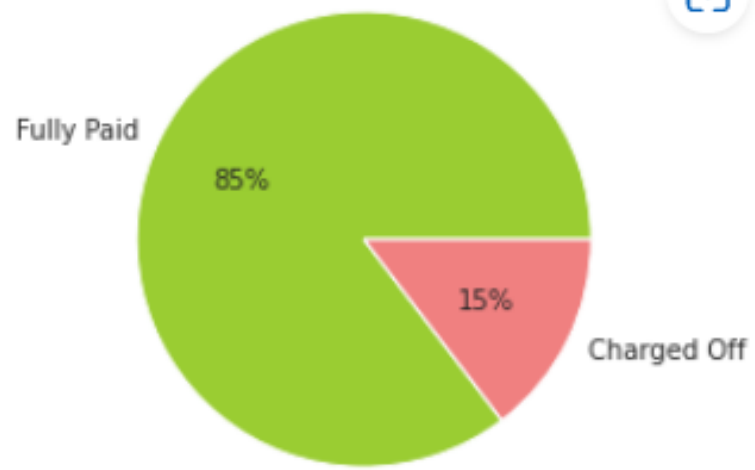

Plot of issue_month
issue_month


Plot of term
36 months
60 months

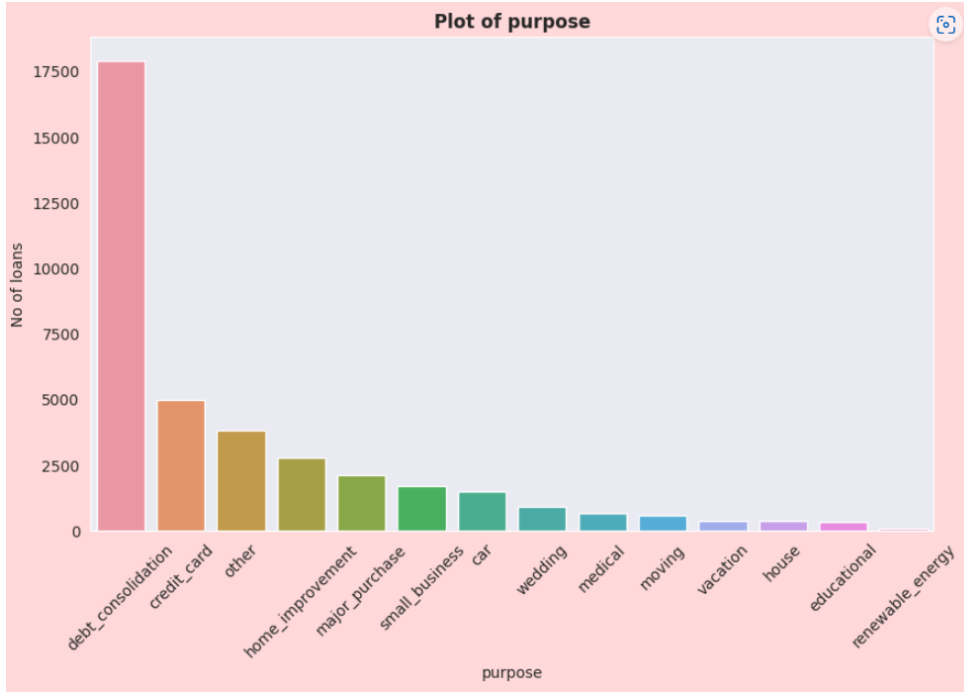

Plot of emp_length
10+ years
< 1 year
2 years
3 years
4 years
5 years
1 year
6 years
7 years
emp_length


Plot of sub_grade
No of loans
sub_grade





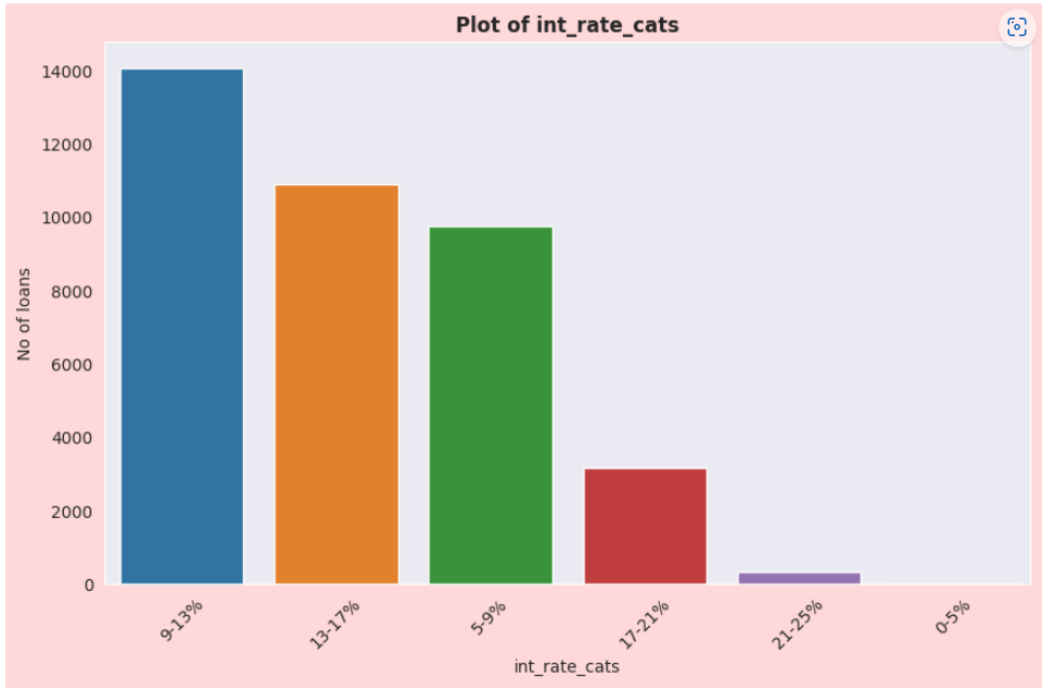
+ Code + Markdown

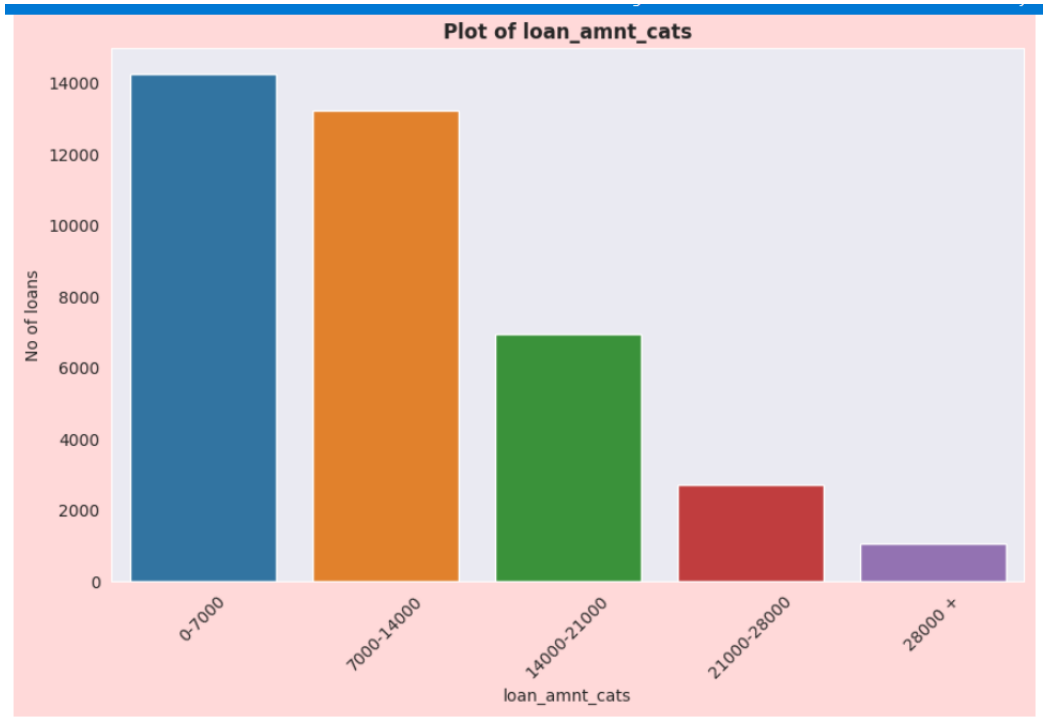


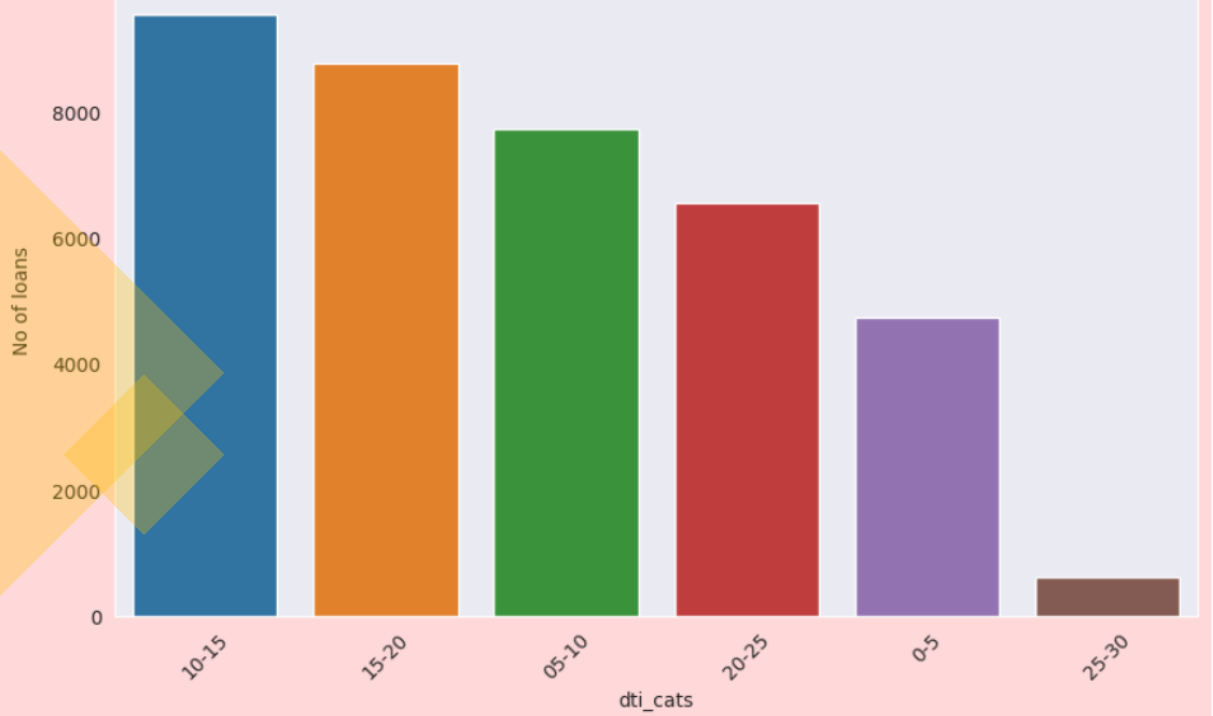
Observations from univariate analysis

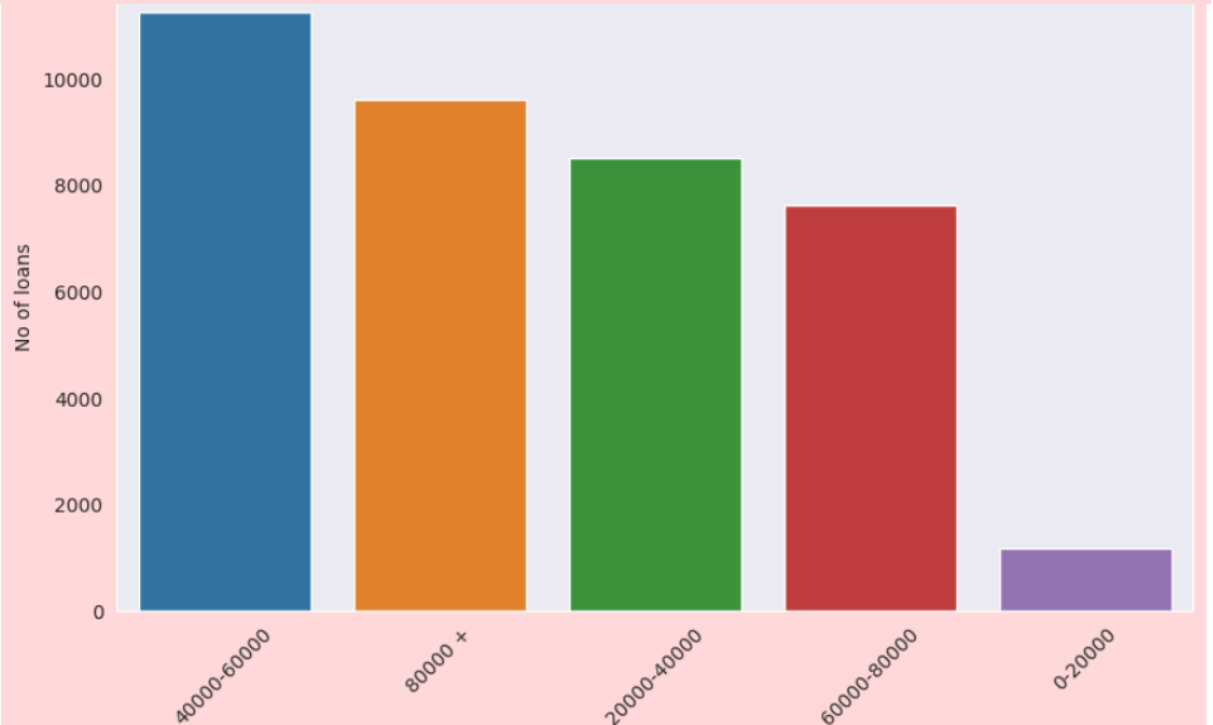
Annual Income had outliers which are far away from the upper fence and treated by removing the values above 99 percentile	Most of the applicants had annual income between 50,000 to 1,00,000	Most loans had interest rate between 12 - 14 %	Most loans had 200 - 400 installments	Most loans had DTI between 10 - 20
There are too many small loans in the range 5,000 to 10,000.	Plot on employment length shows that borrowers with experience more than 10 + years tend to get lot of loans than borrowers with lesser experience	Plot on issue year and month shows that the no. of loans tend to increase with year from 2007 -2011. There are more no. of loans on 2011	Plot on terms shows that 36 month loans are issued more when compared to 60 month loans	Plot on grade shows more loans are given to B, A and C
Plot on sub-grade shows more loans are given to A4, B3, A5, B5, B4	Plot of issue month shows there is increasing trend in number of loans with increase in the months. Maximum loans were given in the month of Oct, Nov, Dec.May be due to festive season	Most of the borrowers dont have a own house. They either live in rented or mortgaged ones.	There is a good trend that shows most of the loans are for debt consolidation and credit card	15 % of the loans are charged off

Segmented Univariate



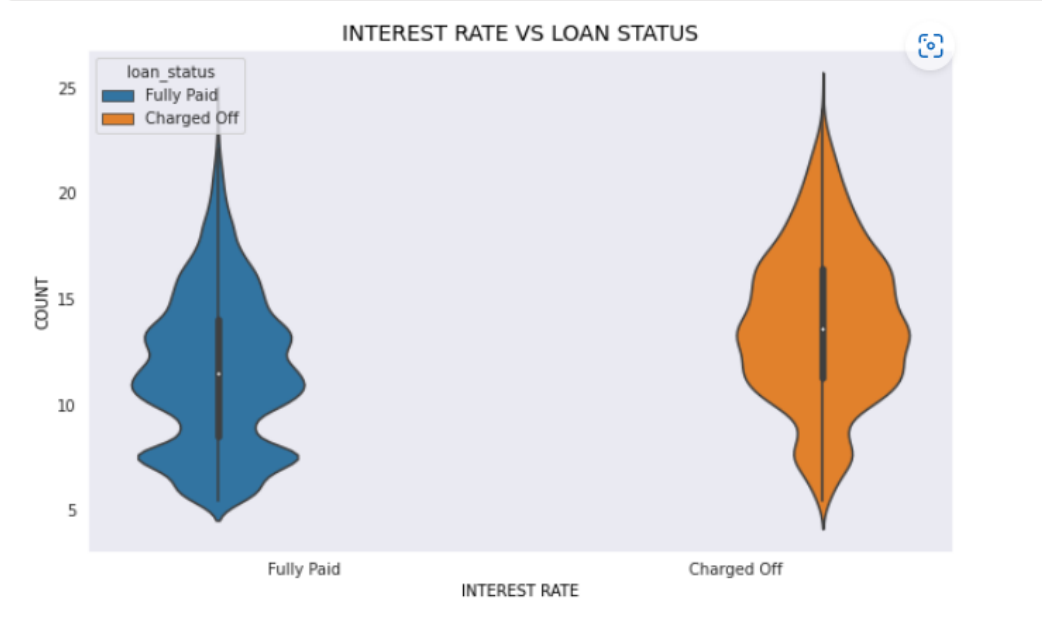


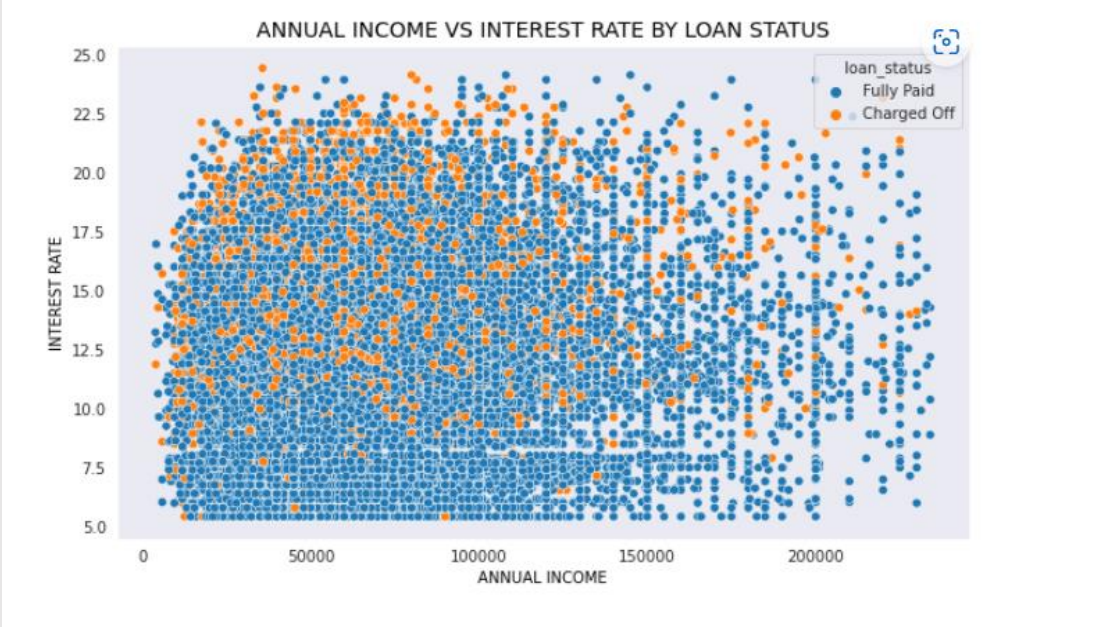


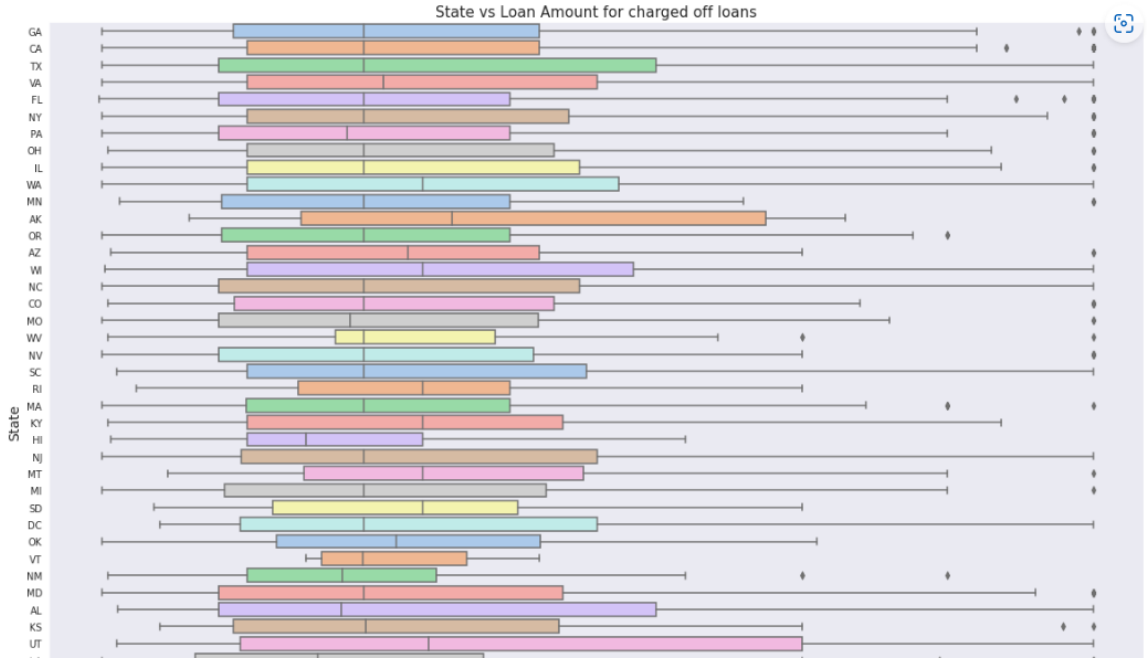


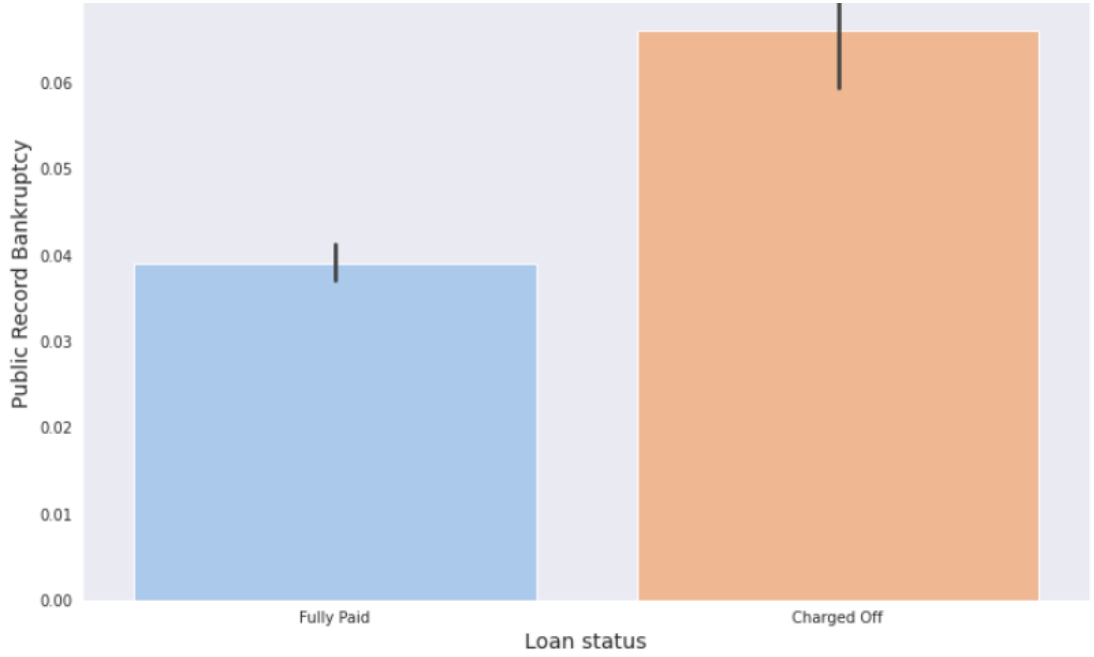
- - Most of the loans have interest rate category of 9-13% and very few loans have interest rate range of 21- 25%
- - Most of the loans has DTI category of 10-15 is
- - Most of the borrowers have annual income range of 40,000 - 60,000
- - Most of the loan amounts is less than 7000

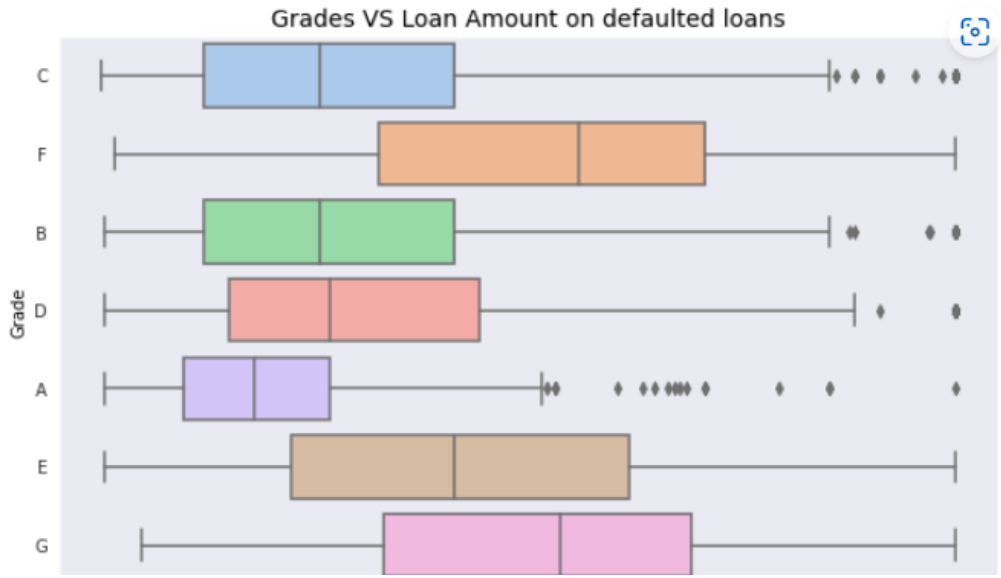
Bivariate Analysis

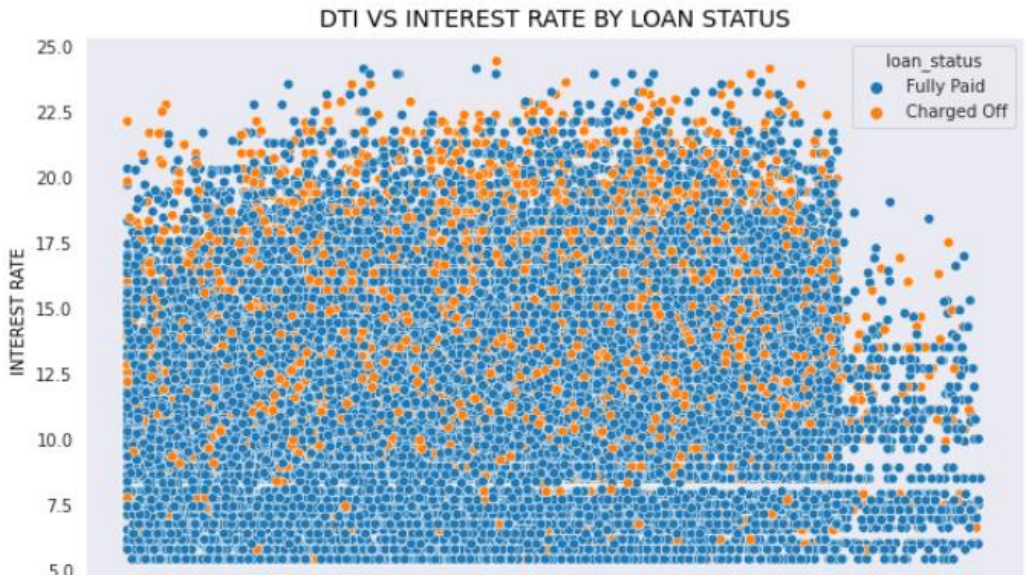


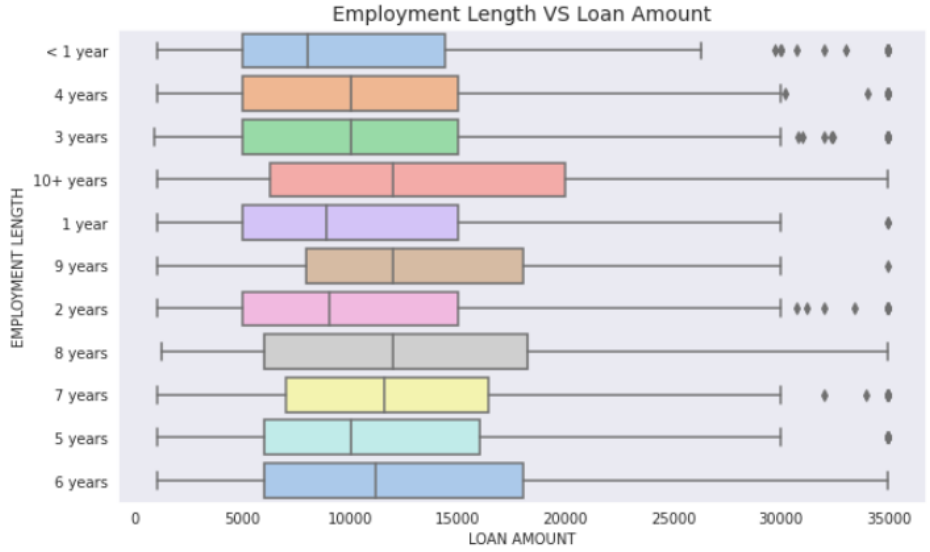


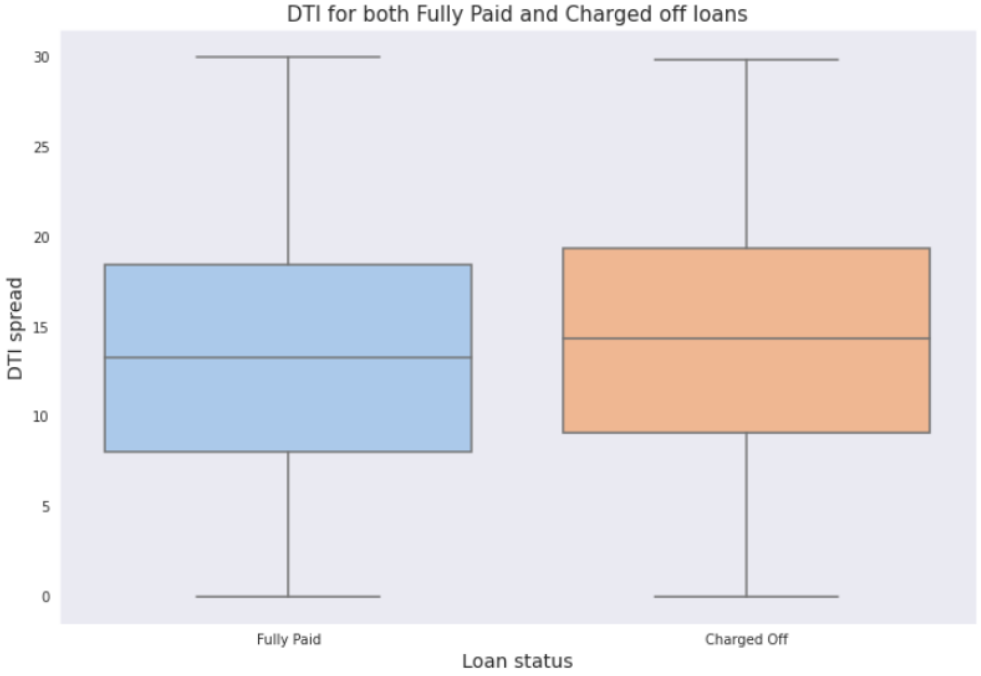

State vs Loan Amount for charged off loans
State
GA
CA
TX
VA
FL
NY
PA
OH
IL
WA
MN
AK
OR
AZ
WI
NC
CO
MO
WV
NV
SC
RI
MA
KY
HI
NJ
MT
MI
SD
DC
OK
VT
NM
MD
AL
KS
UT

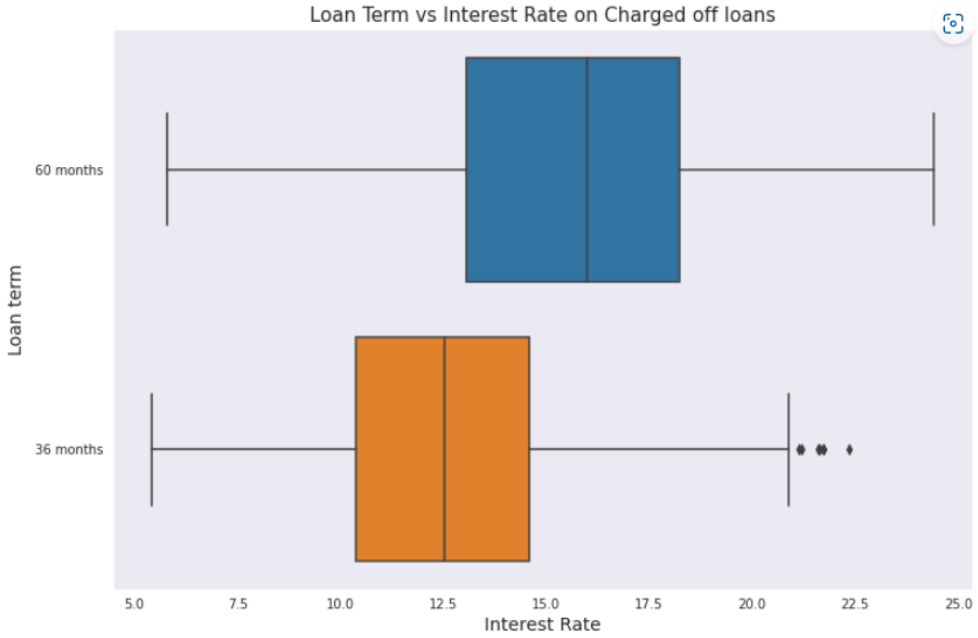

Public Record Bankruptcy
0.06
0.05
0.04
0.03
0.02
0.01
0.00
Fully Paid
Charged Off
Loan status

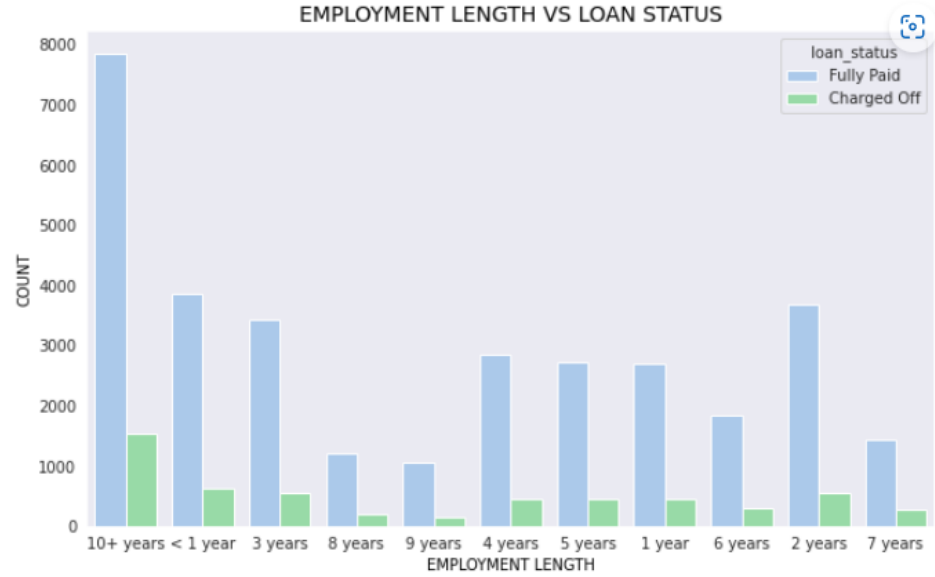

Grades VS Loan Amount on defaulted loans
Grade
C
F
B
D
A
E
G

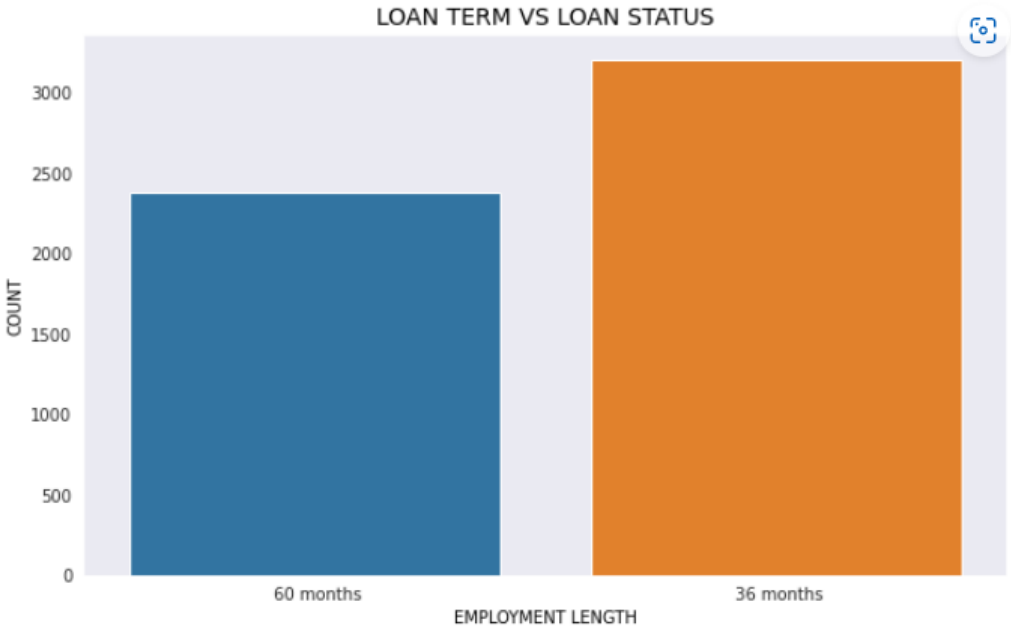

DTI VS INTEREST RATE BY LOAN STATUS
INTEREST RATE
25.0
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0
loan_status
Fully Paid
Charged Off

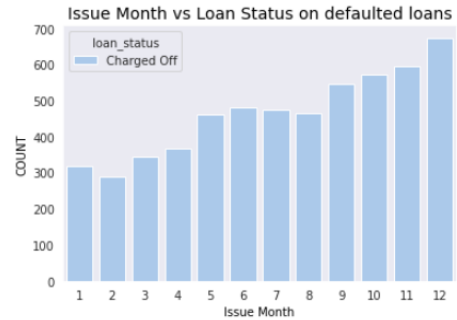

Employment Length VS Loan Amount
< 1 year
4 years
3 years
10+ years
1 year
9 years
2 years
8 years
7 years
5 years
6 years
EMPLOYMENT LENGTH
0
5000
10000
15000
20000
25000
30000
35000
LOAN AMOUNT


DTI for both Fully Paid and Charged off loans
30
25
20
15
10
5
0
DTI spread
Fully Paid
Charged Off
Loan status

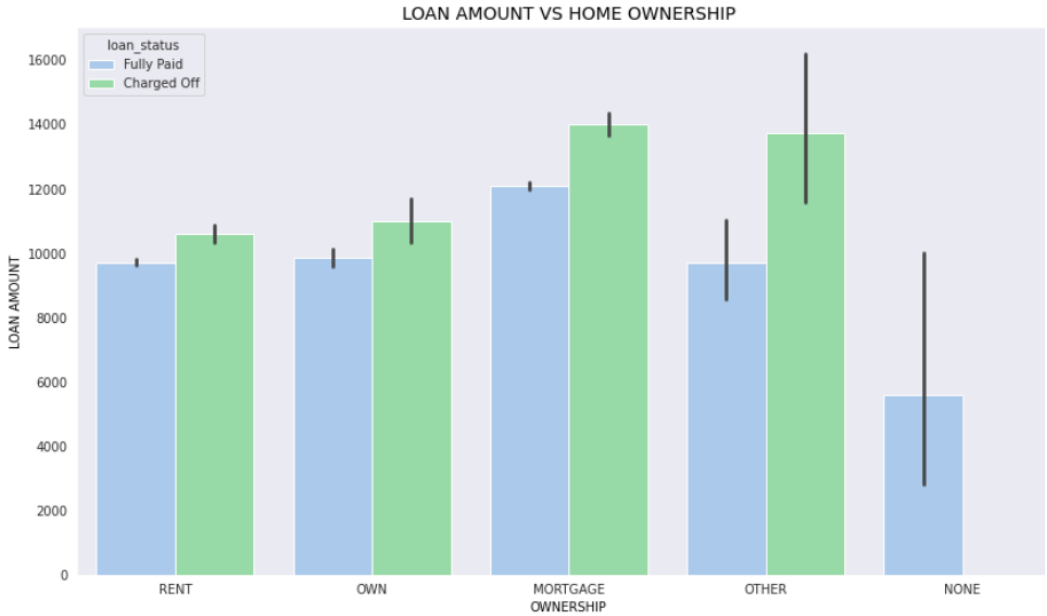

Loan Term vs Interest Rate on Charged off loans
60 months
36 months
Loan term
5.0
7.5
10.0
12.5
15.0
17.5
20.0
22.5
25.0
Interest Rate

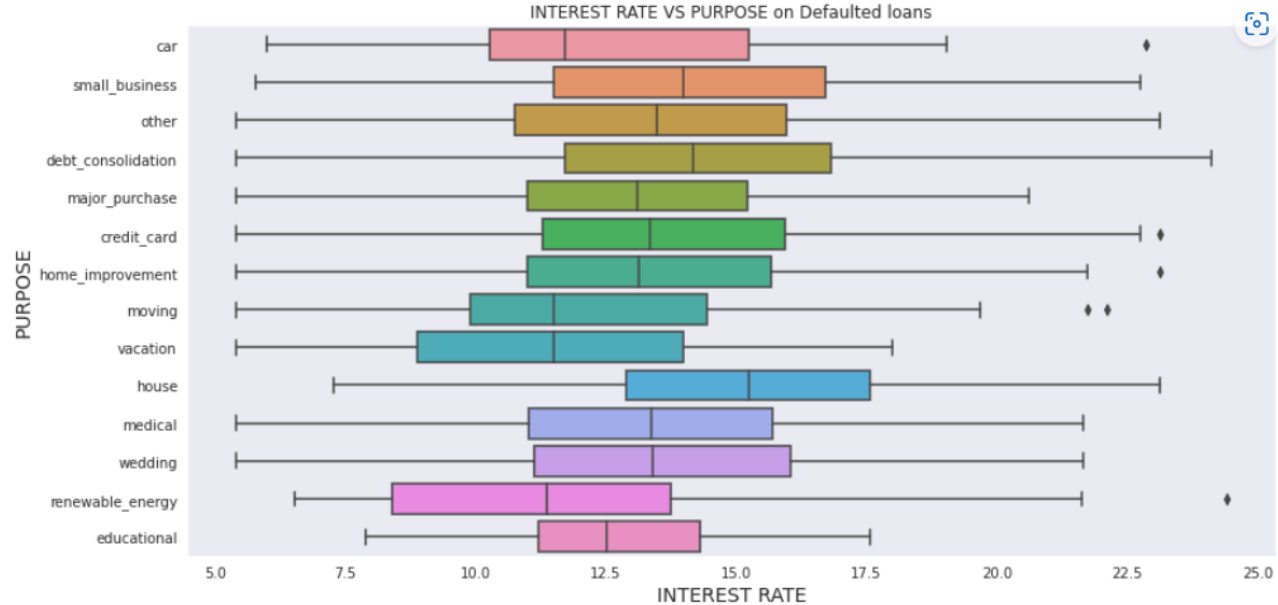

EMPLOYMENT LENGTH VS LOAN STATUS
loan_status
Fully Paid
Charged Off
COUNT
0
1000
2000
3000
4000
5000
6000
7000
8000
10+ years
< 1 year
3 years
8 years
9 years
4 years
5 years
1 year
6 years
2 years
7 years
EMPLOYMENT LENGTH

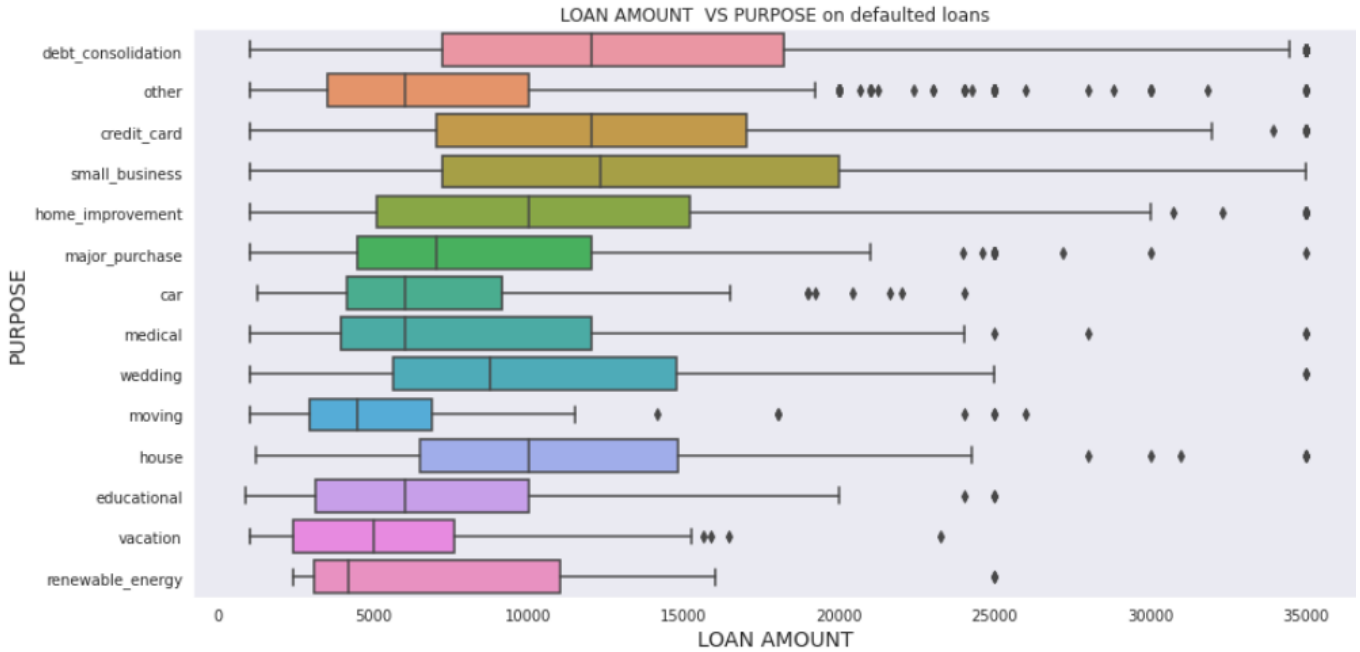

LOAN TERM VS LOAN STATUS
COUNT
0
500
1000
1500
2000
2500
3000
60 months
36 months
EMPLOYMENT LENGTH

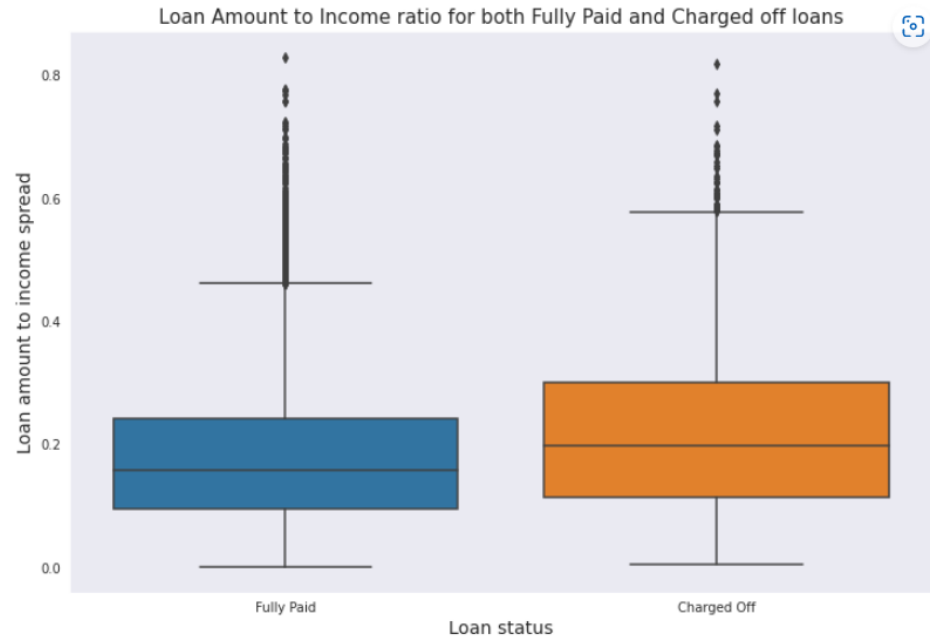

Issue Month vs Loan Status on defaulted loans
loan_status
Charged Off
COUNT
0
100
200
300
400
500
600
700
1
2
3
4
5
6
7
8
9
10
11
12
Issue Month


Issue Year vs Loan Status on defaulted loans
loan_status
Charged Off
COUNT
0
500
1000
1500
2000
2500
3000
2007
2008
2009
2010
2011
Issue Year


LOAN AMOUNT VS HOME OWNERSHIP
loan_status
Fully Paid
Charged Off
LOAN AMOUNT
OWNERSHIP
RENT
OWN
MORTGAGE
OTHER
NONE


INTEREST RATE VS PURPOSE on Defaulted loans
PURPOSE
car
small_business
other
debt_consolidation
major_purchase
credit_card
home_improvement
moving
vacation
house
medical
wedding
renewable_energy
educational
INTEREST RATE


LOAN AMOUNT VS PURPOSE on defaulted loans
PURPOSE
debt_consolidation
other
credit_card
small_business
home_improvement
major_purchase
car
medical
wedding
moving
house
educational
vacation
renewable_energy
LOAN AMOUNT


Loan Amount to Income ratio for both Fully Paid and Charged off loans
Loan amount to income spread
Fully Paid
Charged Off
Loan status

Bivariate Analysis - Observations

The violin plot shows the distribution of interest rates on charged off loans on the higher end 13 - 22% when compared to fully-paid loans, which reveals interest rate is one of the driving factor for loan default

When the loan interest rate is greater than 16 % and the annual income of the person is less than 100000, the loan is more likely to be defaulted. This reveals annual income and interest rate has a positive correlation to loan default.

When the debt to income ratio is more than 13 and interest rate is higher than 15 likely to default

Giving loans at interest rate when the annual income is < 1,50,000 leads to loan default

Customers with DTI greater than 15 is more likely to default

Grades F, G and H are the top contributors of loan defaults

Grade F and G have the median around 20k and Q3 at 25k

Grade A is the least with median at 7.5k

Most of states have median is at 10k for charged off loan

State AK, MT, KY, UT, TN are marginally above the 10k level and at 13k

States AK, UT the Q3 is at 23k

DTI does not seem to be contributing because the median is close to each other for Fully Paid and Charged Off

36 months term loans tend to default more

Employees with experience more than 10+ years contribute to loan defaults more followed by year 6, 7, 8 and 9 .

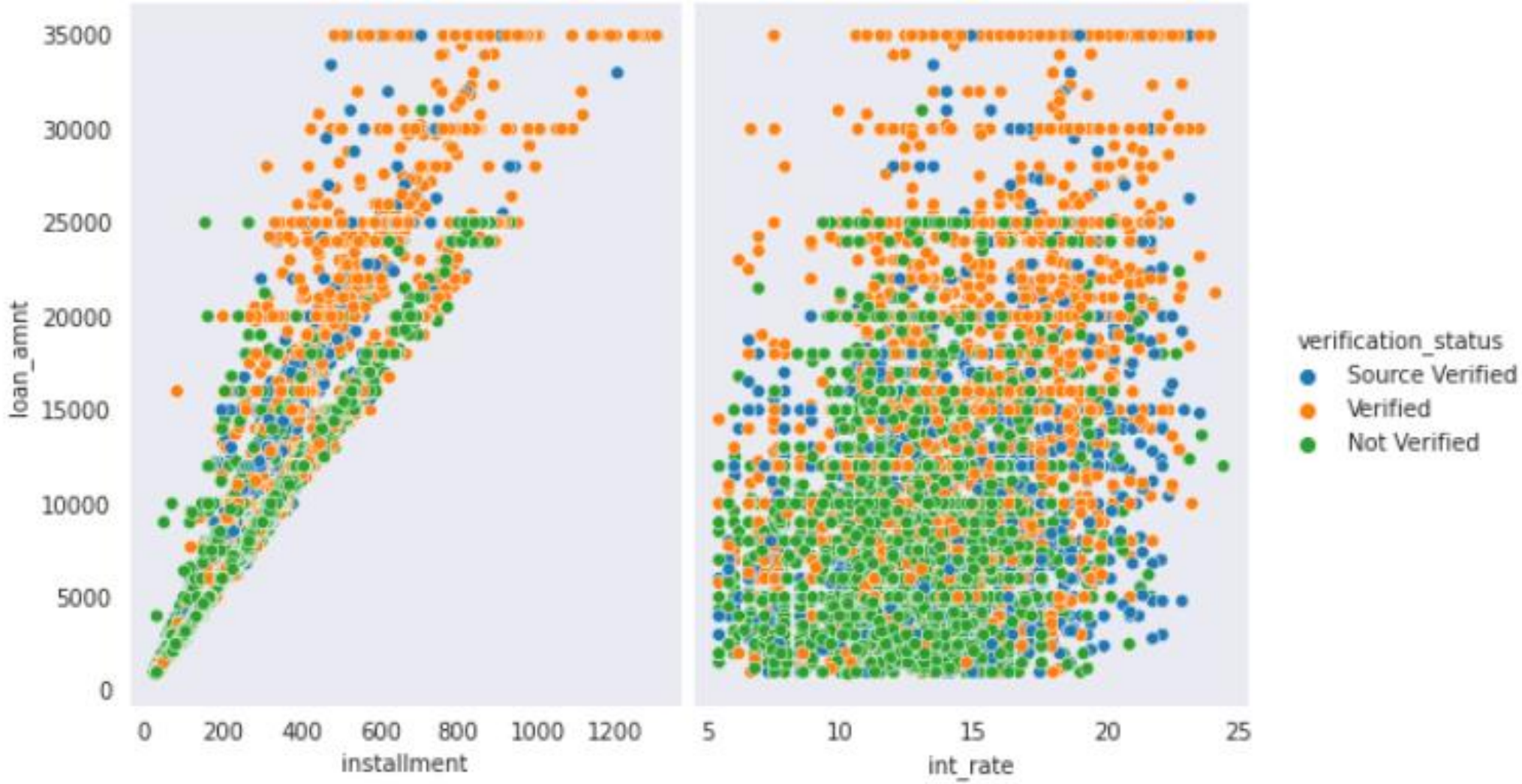
Borrowers with mortgaged home with loan amount greater than 14,000 tend to default more

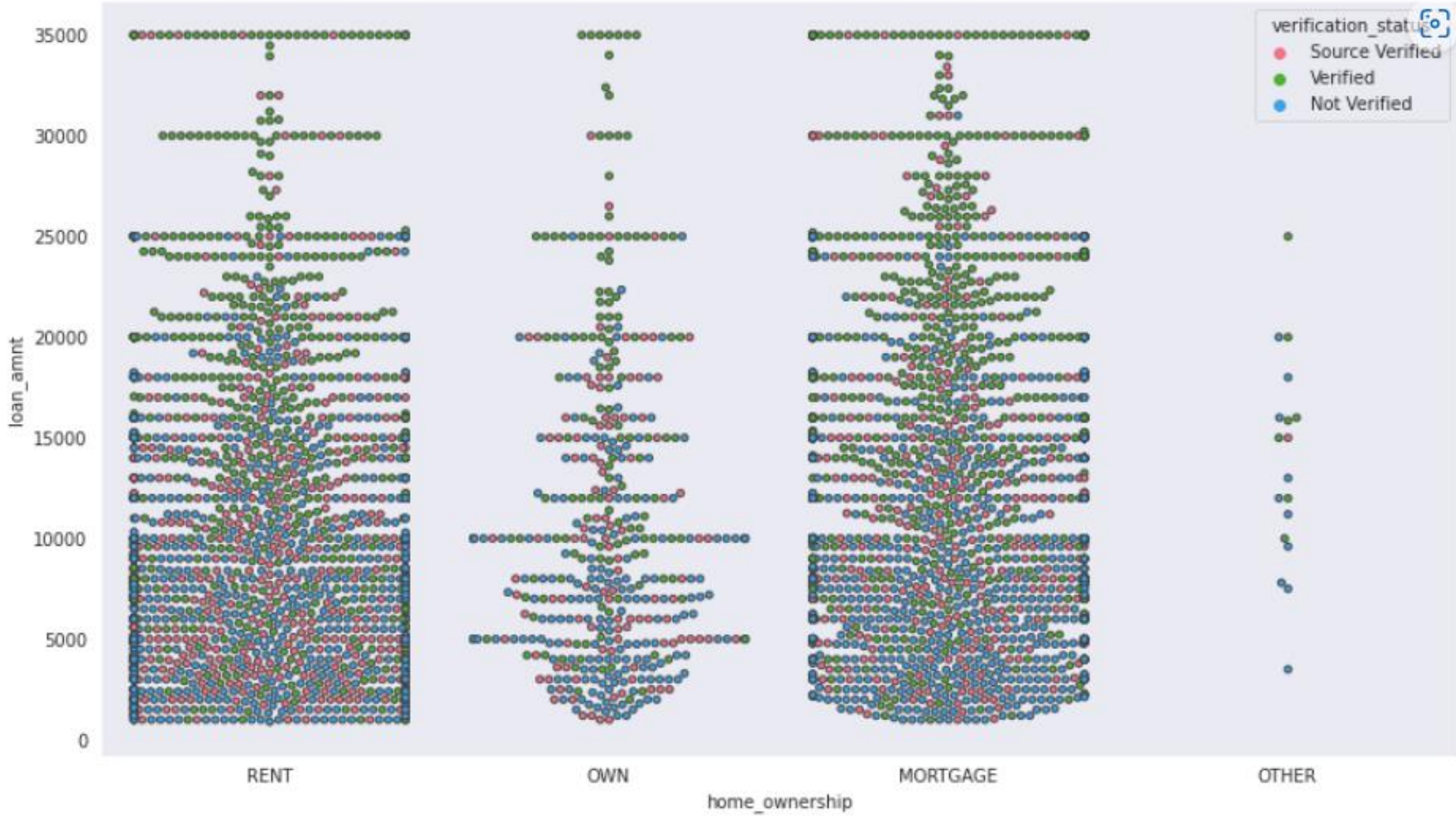
Housing loans with higher interest rate charge-off more

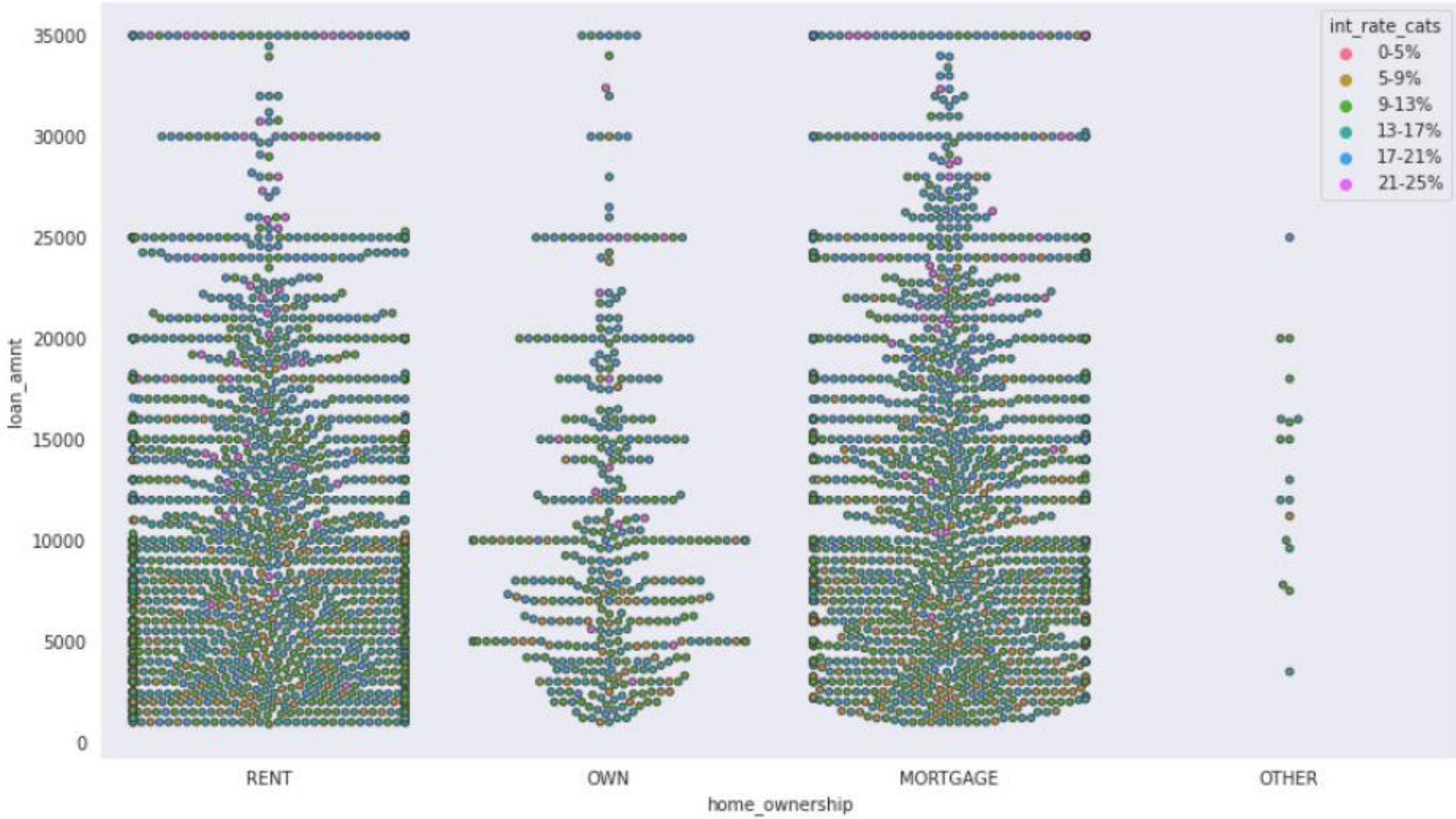
Debt consolidation, credit card and small business are the top loan purpose that contribute to defaults

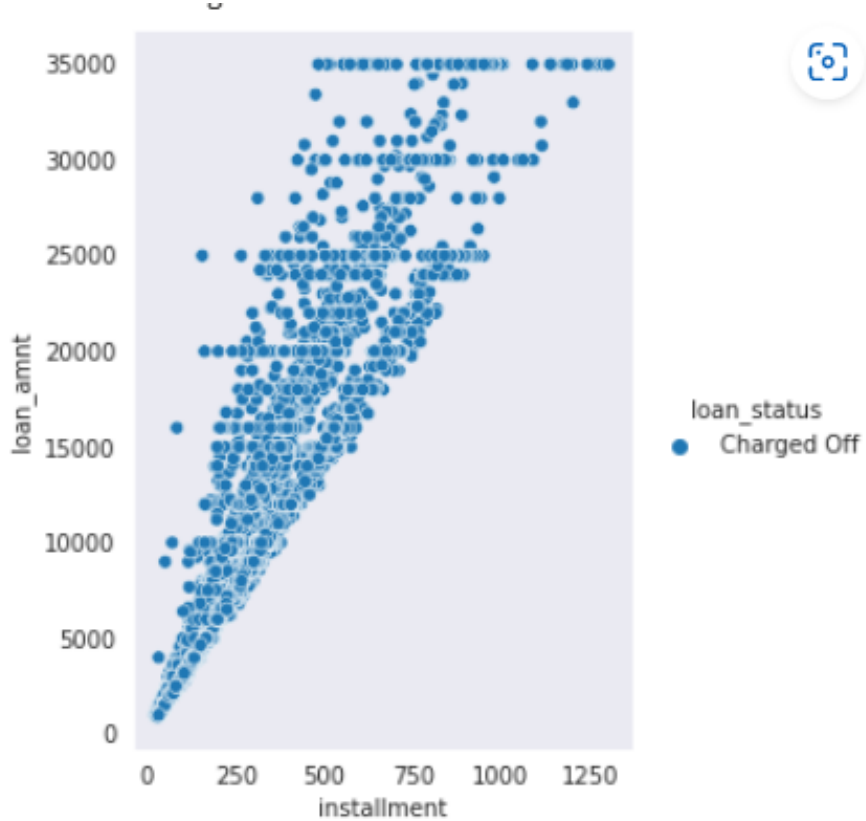
Loan amount to income ratio contributes to loan defaults

Multivariate Analysis


35000
30000
25000
20000
15000
10000
5000
0
loan_amnt
0
200
400
600
800
1000
1200
installment
5
10
15
20
25
int_rate
verification_status
Source Verified
Verified
Not Verified


verification_status
Source Verified
Verified
Not Verified
35000
30000
25000
20000
15000
10000
5000
0
loan_amnt
RENT
OWN
MORTGAGE
OTHER
home_ownership


int_rate_cats
0-5%
5-9%
9-13%
13-17%
17-21%
21-25%
loan_amnt
35000
30000
25000
20000
15000
10000
5000
0
RENT
OWN
MORTGAGE
OTHER
home_ownership


35000
30000
25000
20000
15000
10000
5000
0
loan_amnt
0
250
500
750
1000
1250
installment
loan_status
Charged Off

Driving Factors for Loan Default

Higher interest rate (13-22%)	Employees with experience more than 10+ years contribute to loan defaults more followed by year 6, 7, 8 and 9 .	Interest rate greater than 16 % and the annual income of the person is less than 100000	DTI ratio more than 13 and interest rate is higher than 15%	Loan defaulting increases with increase in public bankruptcy records.	Borrowers with mortgaged home and rented home default more than people with own home.
Housing loans with higher interest rate (above 9 %)	Loan amount to income ratio contributes to loan defaults	Grades F, G and H are the top contributors of loan defaults	Shorter loan tenure (36 months)	State AK, MT, KY, UT, TN borrowed marginally above the 10k level median 13k and contribute to defaults.	Debt consolidation, credit card and small business are the top loan purpose that contribute to defaults
	Loan amount to income ratio higher than .5 contributes to loan defaults	Verified loan amounts are given to employee with good experience and they tend to default.	Verified loans with high loan amount for borrowers with mortgage and rental homes	Installments increase with higher loan amount and so is the defaulting rate	

Reccomendations

Avoid giving high amount loans for people without own house. The loan amount can be limited to 15,000 with less interest rate (< 9 %) to avoid risk of default.

Avoid giving higher amount loans to employees with 10+ years of experience with higher interest rate (> 13 %) .

Verifying the employment status should not result in giving higher loan amount. Other factors like DTI, Loan amount to income ratio, Public bankrupt records should be considered

Be extra cautious for loans taken for debt consolidation, credit cards especially during the October, November, December.