

LENDING CLUB CASE STUDY

AGENDA

PROBLEM SUMMARY

1

**DATA ONBOARDING
& INITIAL EDA**

2

**DATA PREPROCESSING
AND CLEANING**

3



7

INSIGHTS

6

**BIVARIATE ANALYSIS :
LOAN STATUS VS
DIFFERENT CATEGORICAL
VARIABLE**

5

**BIVARIATE ANALYSIS :
LOAN STATUS VS
DIFFERENT NUMERICAL
VARIABLE**

4

**UNIVARIATE ANALYSIS :
NUMERICAL &
CATEGORICAL
VARIABLE**

Problem statement :

A Consumer Finance company which specializes in lending various types of loans to urban customer . We need to analyse and build insight which can be considered by the loan approving team to decide on approval mechanism . Historical loan dataset is available to study the patterns and generate interesting actionable insight

Constraints :

1. If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
2. 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

Insight type :

1. Types of customer / lead demographic to consider for loan approval
2. Leading indicator for loan default / customer behaviour to improve collection management system
3. Whitespaces for product development
4. Types of loan products that are doing well and can be doubled down for market expansion

STEPS : DATA ONBOARDING AND INITIAL EDA



Import all necessary libraries for analysis , mathematical operations and data visualization



Upload the loan dataset from local system after downloading



Initial EDA of the uploaded dataset

STEPS : DATA PREPROCESSING AND CLEANING



Identify and remove columns that has only null values



Remove insignificant columns



Remove columns that have constant values



Remove columns that has high percentage of null columns , consider higher than 25%



Remove rows for all the other columns sequentially that has lower percentage of null values , and we don't have a definite logic for imputation

UNIVARIATE ANALYSIS NUMERICAL & CATEGORICAL VARIABLE

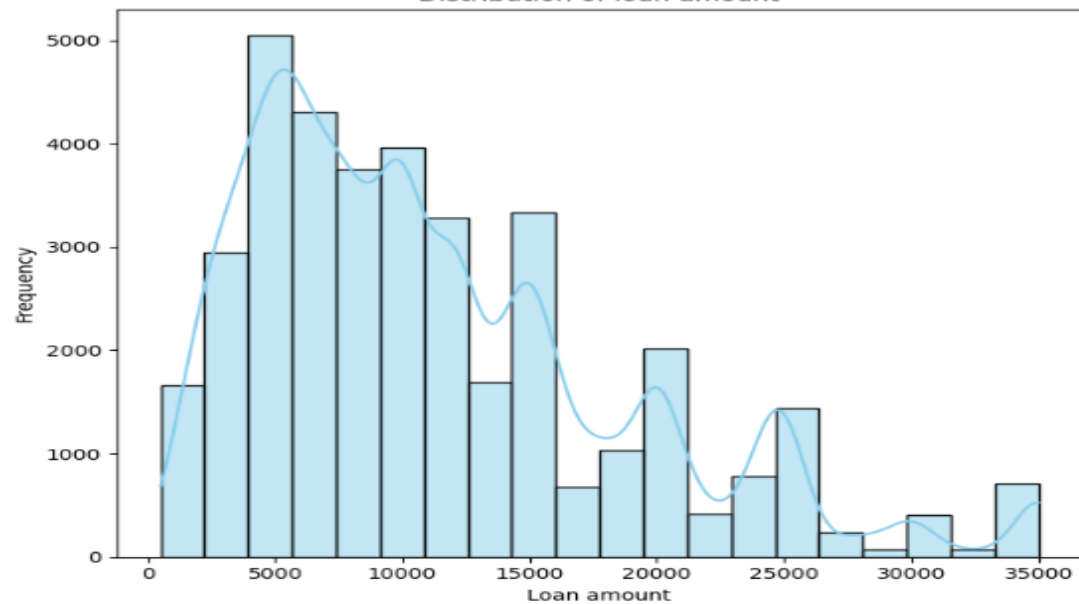
Numerical values

```
Skewness: loan_amnt    1.050861
Kurtosis: loan_amnt    0.743797
Skewness: funded_amnt  1.050861
Kurtosis: funded_amnt  0.743797
Skewness: funded_amnt_inv 1.050861
Kurtosis: funded_amnt_inv 0.743797
Skewness: int_rate     1.050861
Kurtosis: int_rate     0.743797
Skewness: installment  1.050861
Kurtosis: installment  0.743797
Skewness: total_acc    1.050861
Kurtosis: total_acc    0.743797
Skewness: total_pymnt  1.050861
Kurtosis: total_pymnt  0.743797
Skewness: total_pymnt_inv 1.050861
Kurtosis: total_pymnt_inv 0.743797
Skewness: total_rec_prncp 1.050861
Kurtosis: total_rec_prncp 0.743797
Skewness: total_rec_int 1.050861
Kurtosis: total_rec_int 0.743797
Skewness: last_pymnt_amnt 1.050861
Kurtosis: last_pymnt_amnt 0.743797
Skewness: annual_inc   1.050861
Kurtosis: annual_inc   0.743797
Skewness: dti          1.050861
Kurtosis: dti          0.743797
Skewness: delinq_2yrs  1.050861
Kurtosis: delinq_2yrs  0.743797
Skewness: inq_last_6mths 1.050861
Kurtosis: inq_last_6mths 0.743797
Skewness: open_acc     1.050861
Kurtosis: open_acc     0.743797
Skewness: pub_rec      1.050861
Kurtosis: pub_rec      0.743797
Skewness: revol_bal    1.050861
Kurtosis: revol_bal    0.743797
Skewness: revol_util   1.050861
Kurtosis: revol_util   0.743797
```

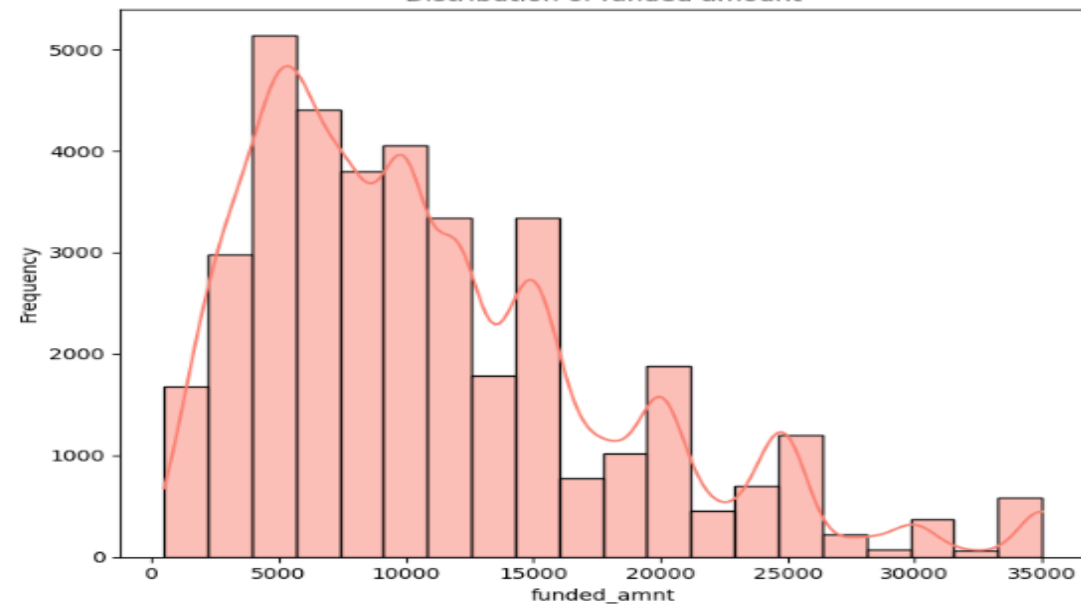
Inference :

1. All the numerical variables considered has Skewness value of around 1 which is an indicator of low right skewed distribution
2. Kurtosis score for all the numerical variable is lesser than 1 which means they are “Platykurtic” , means they have thinner tails with fewer extreme values (outliers)

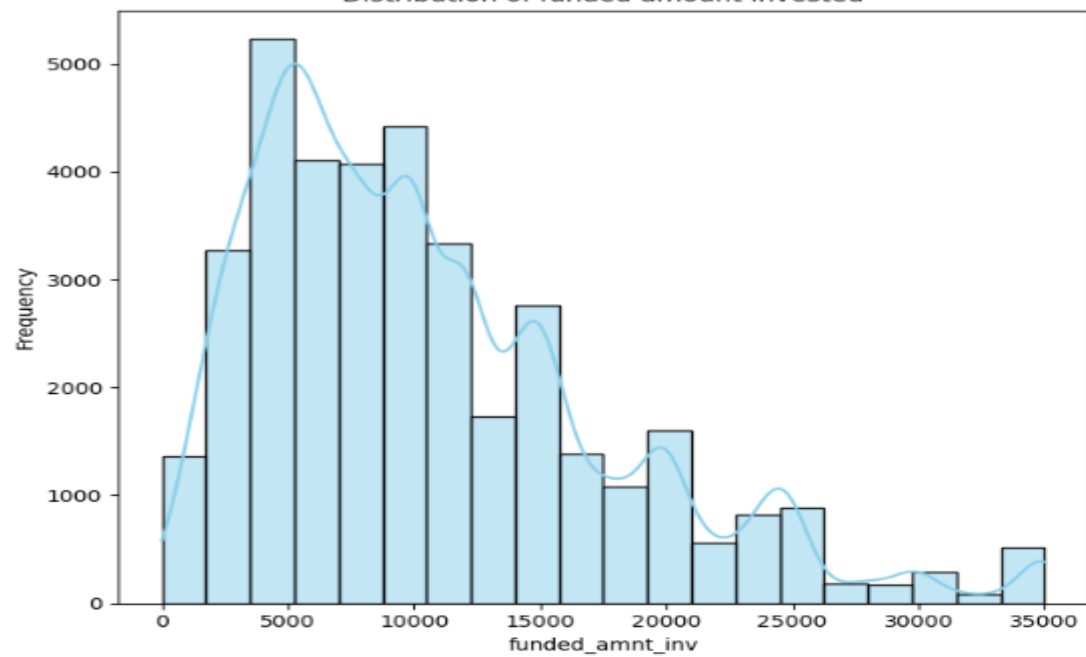
Distribution of loan amount



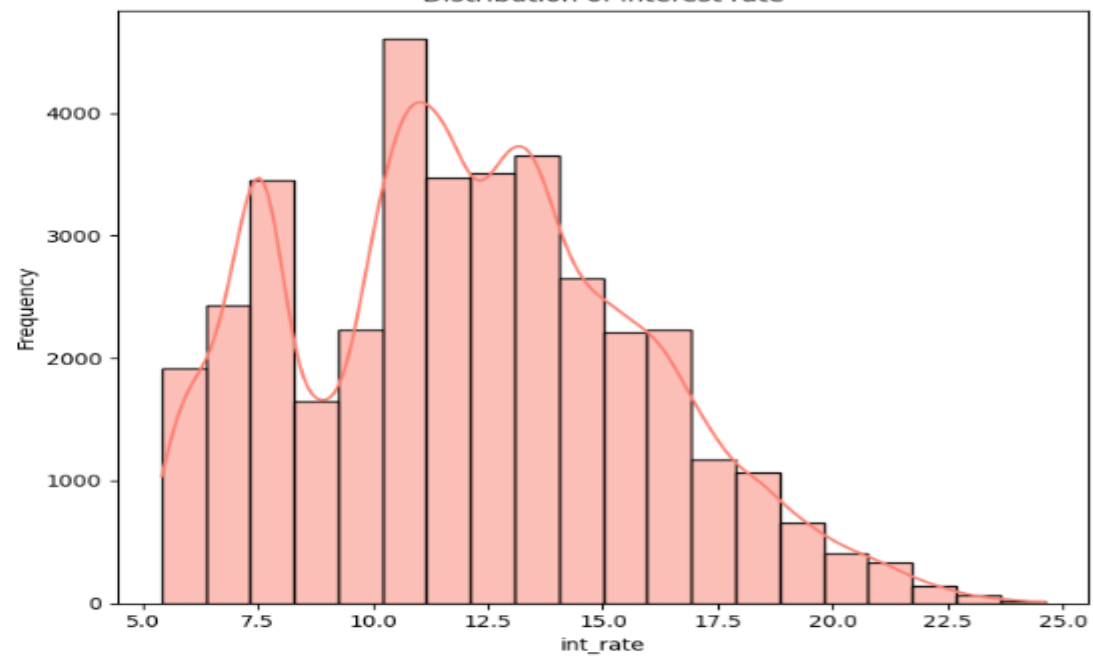
Distribution of funded amount



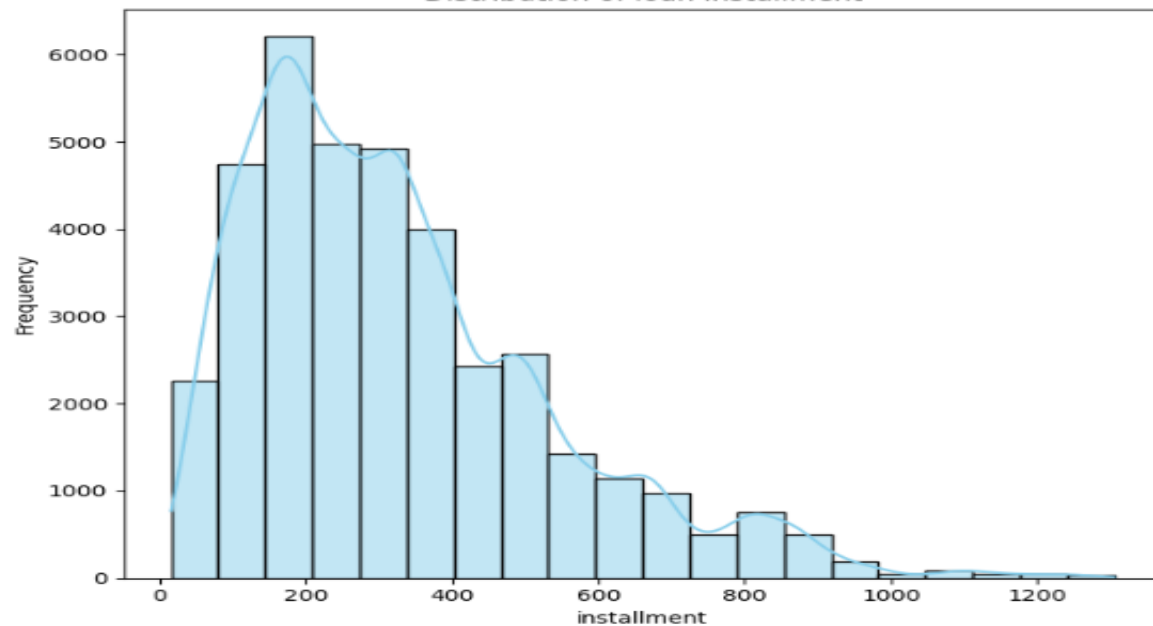
Distribution of funded amount invested



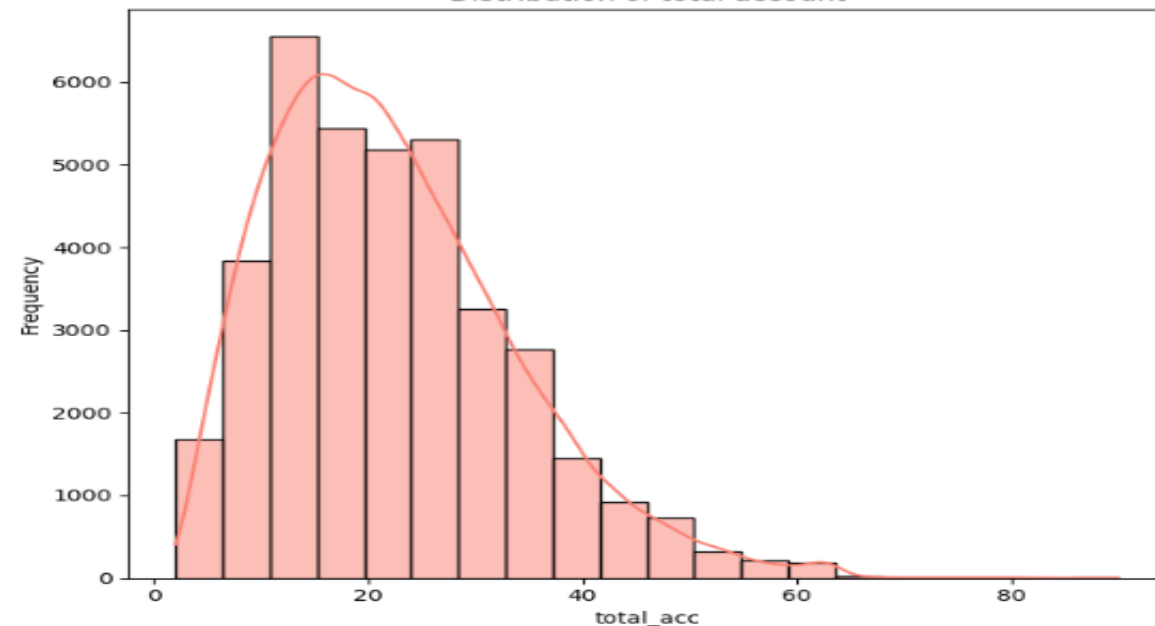
Distribution of interest rate



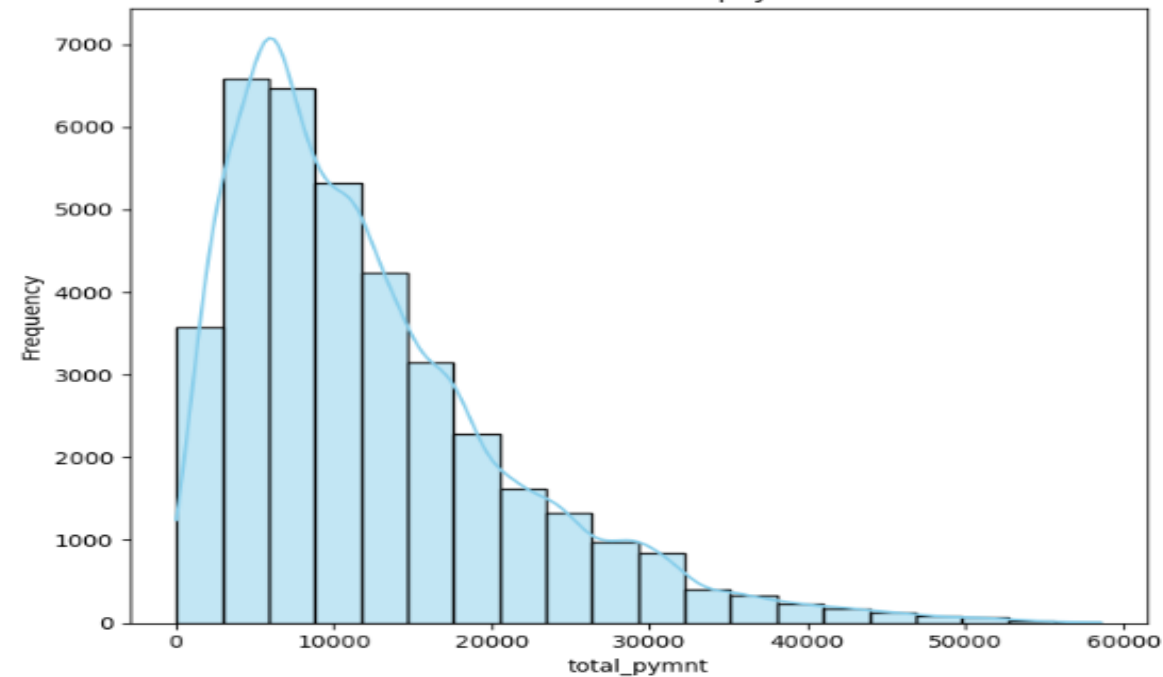
Distribution of loan installment



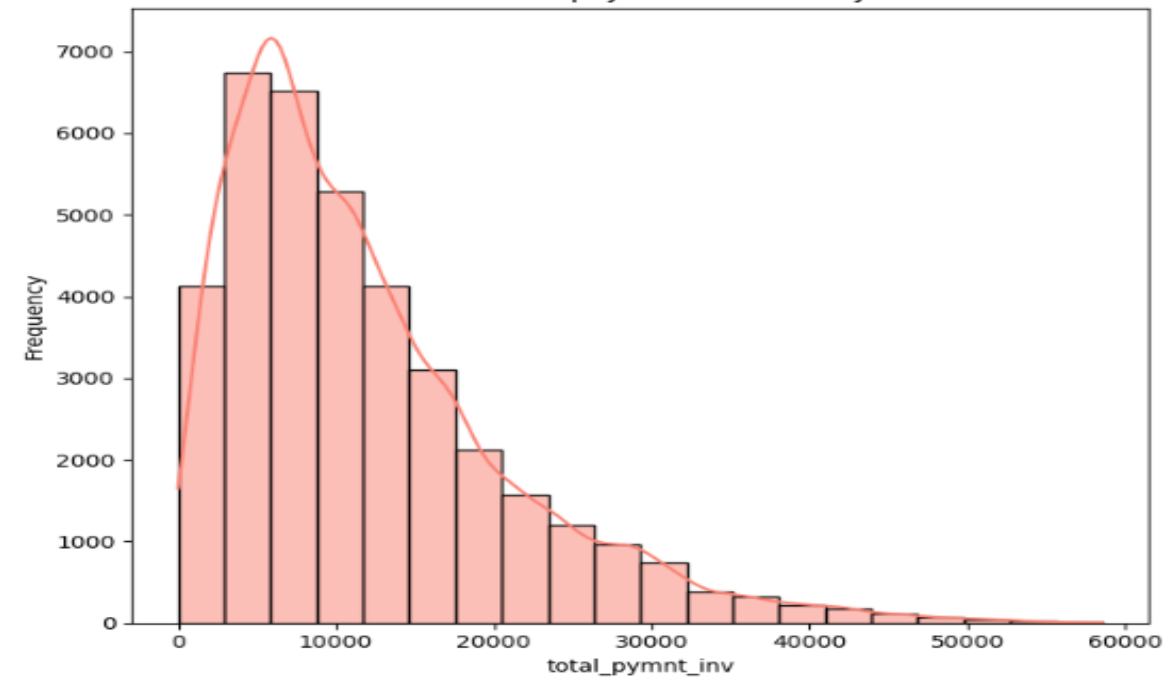
Distribution of total account

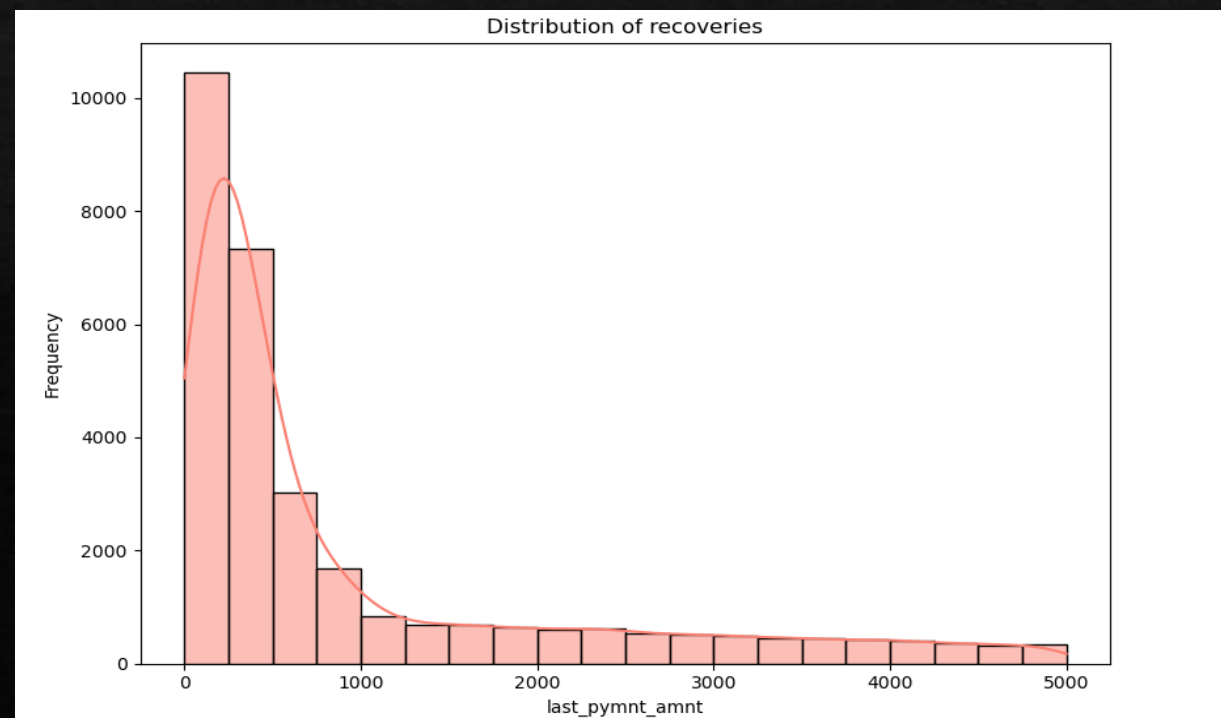
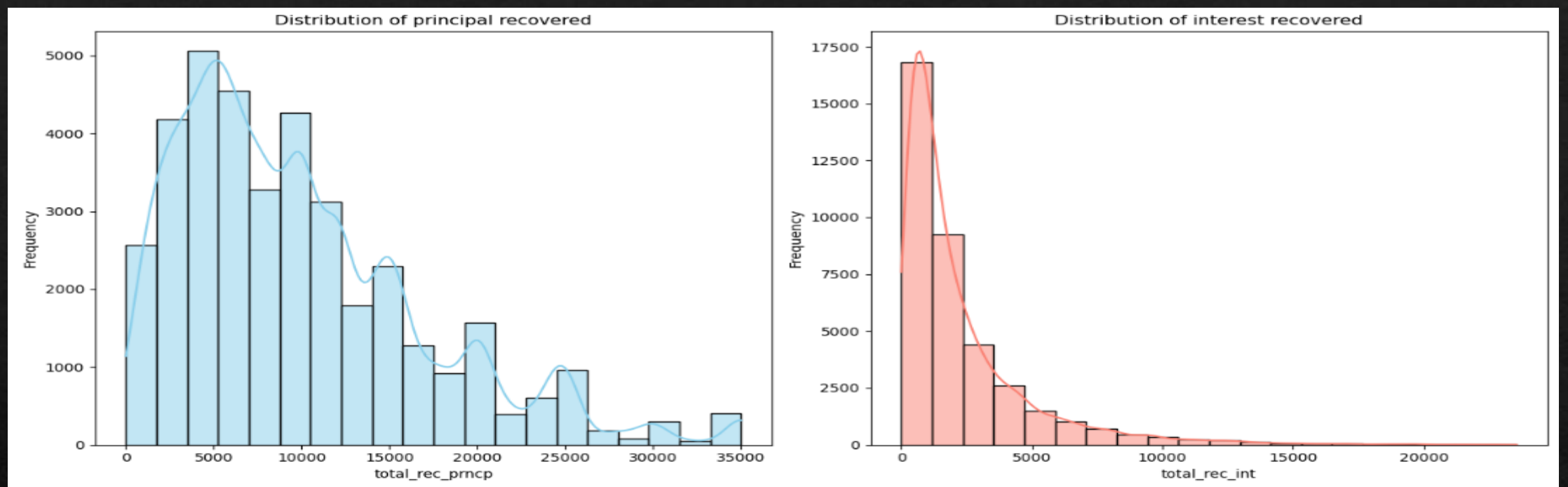


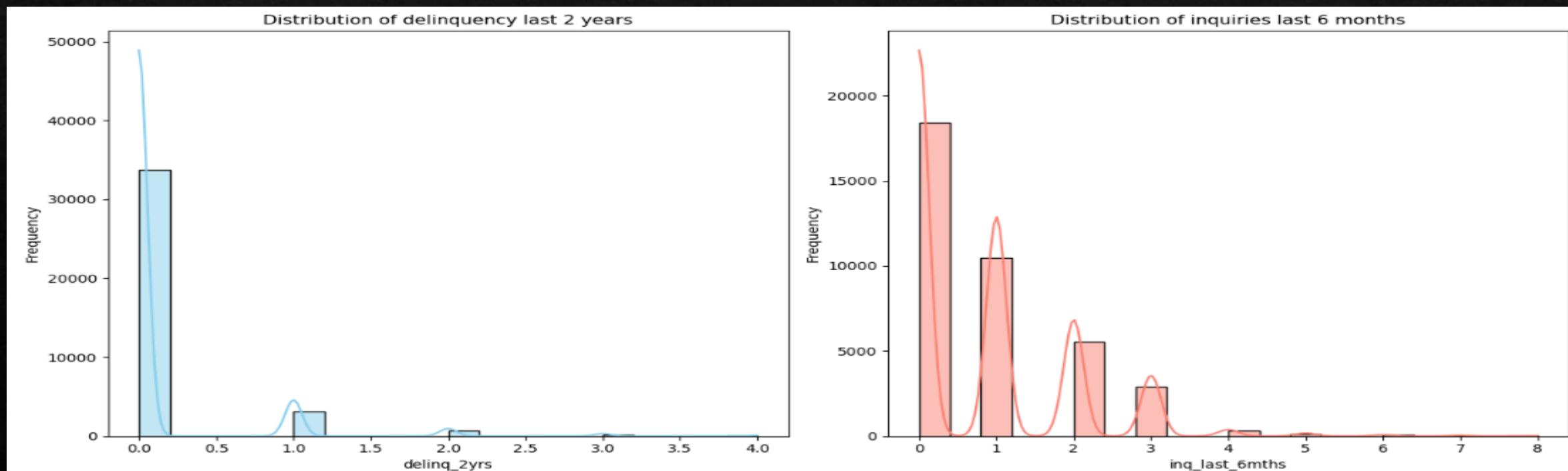
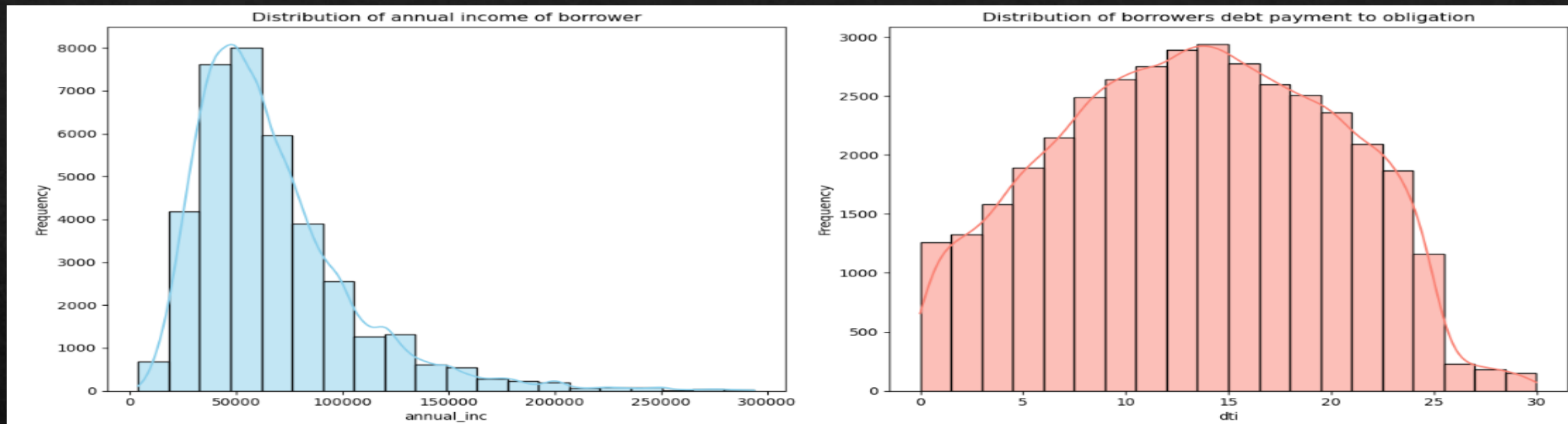
Distribution of total payment



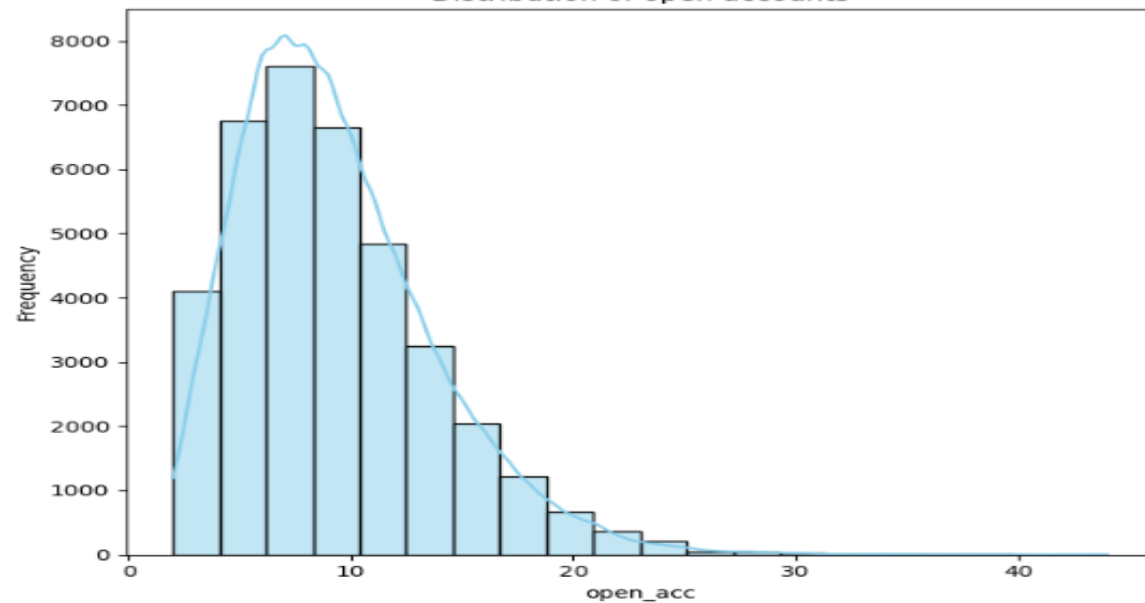
Distribution of total payment received by investor



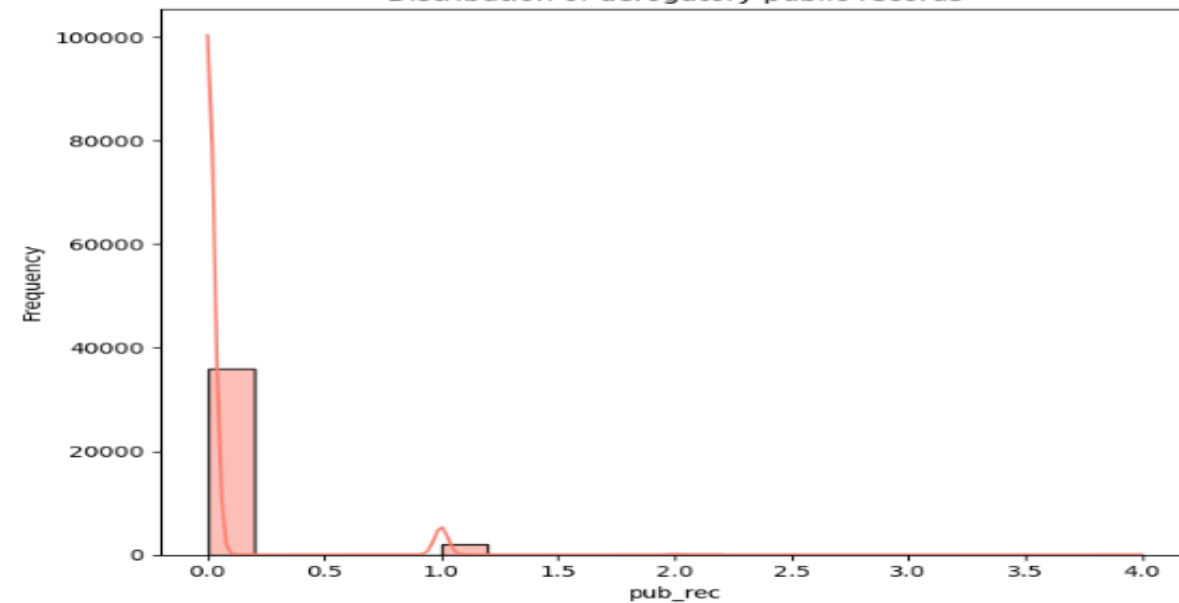




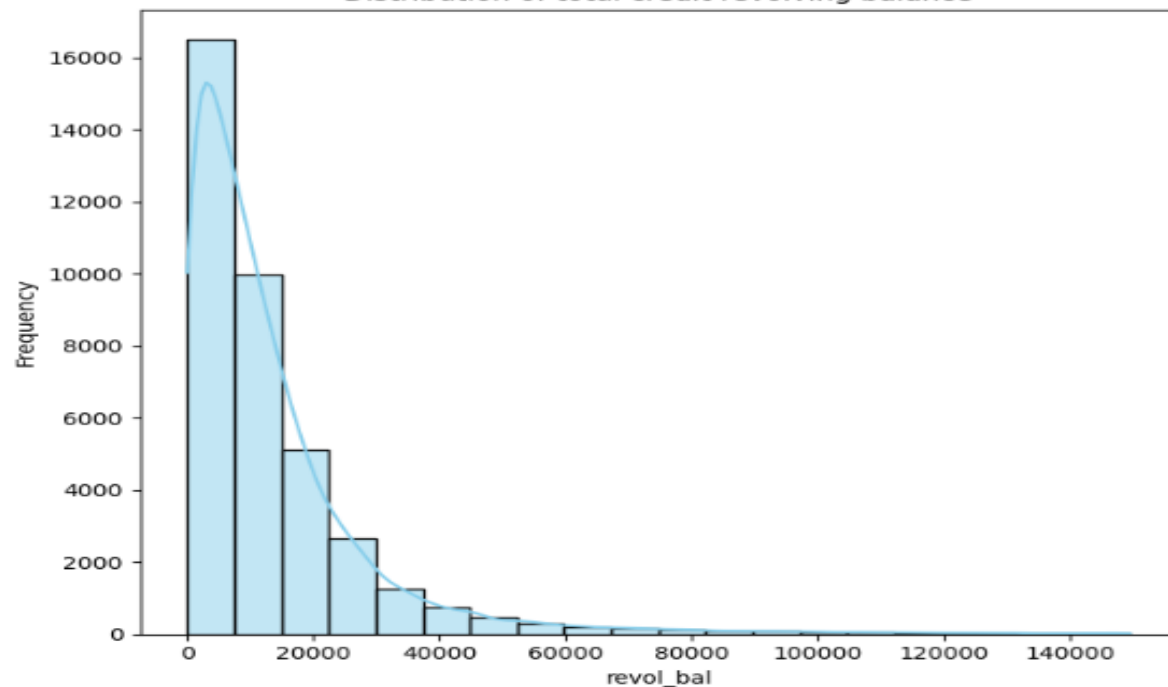
Distribution of open accounts



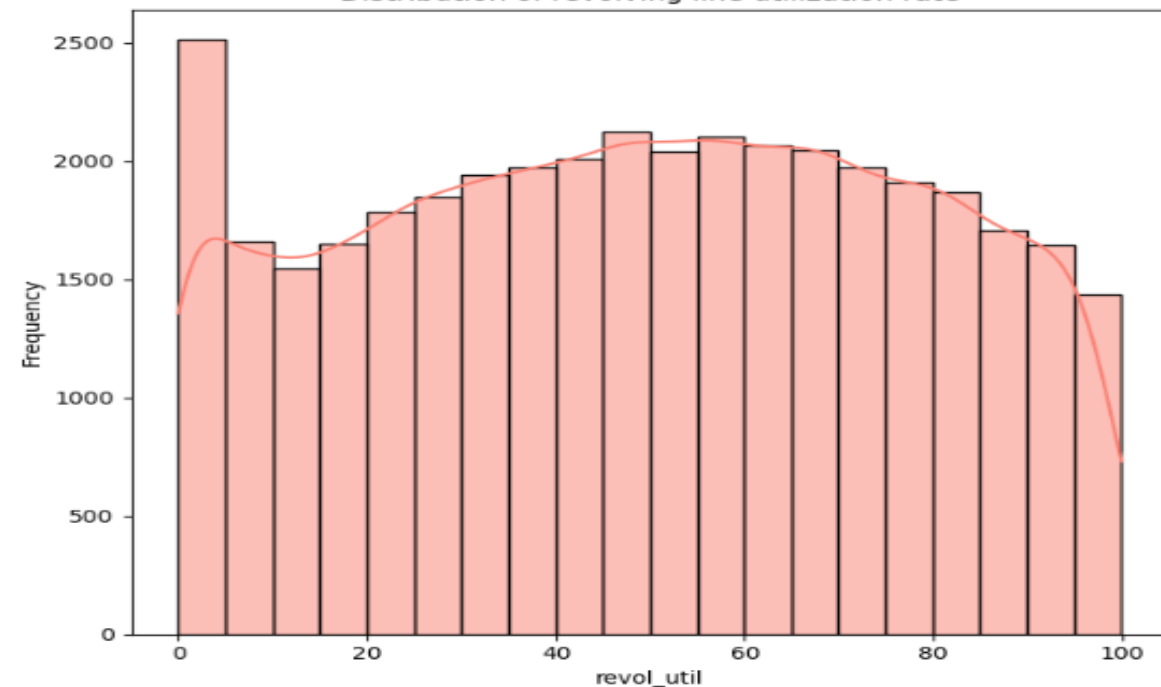
Distribution of derogatory public records



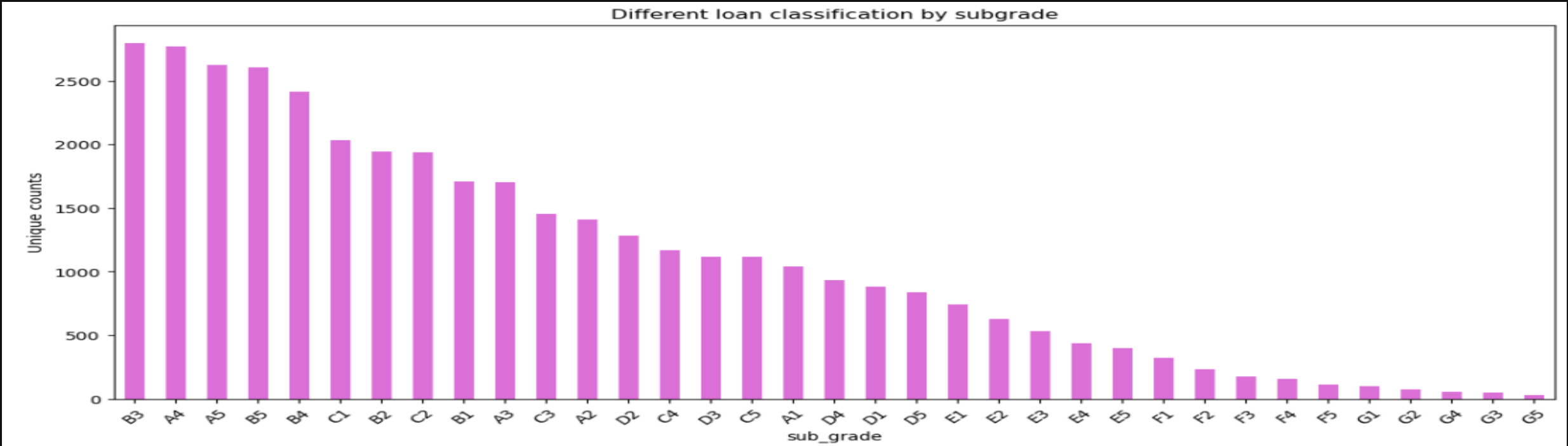
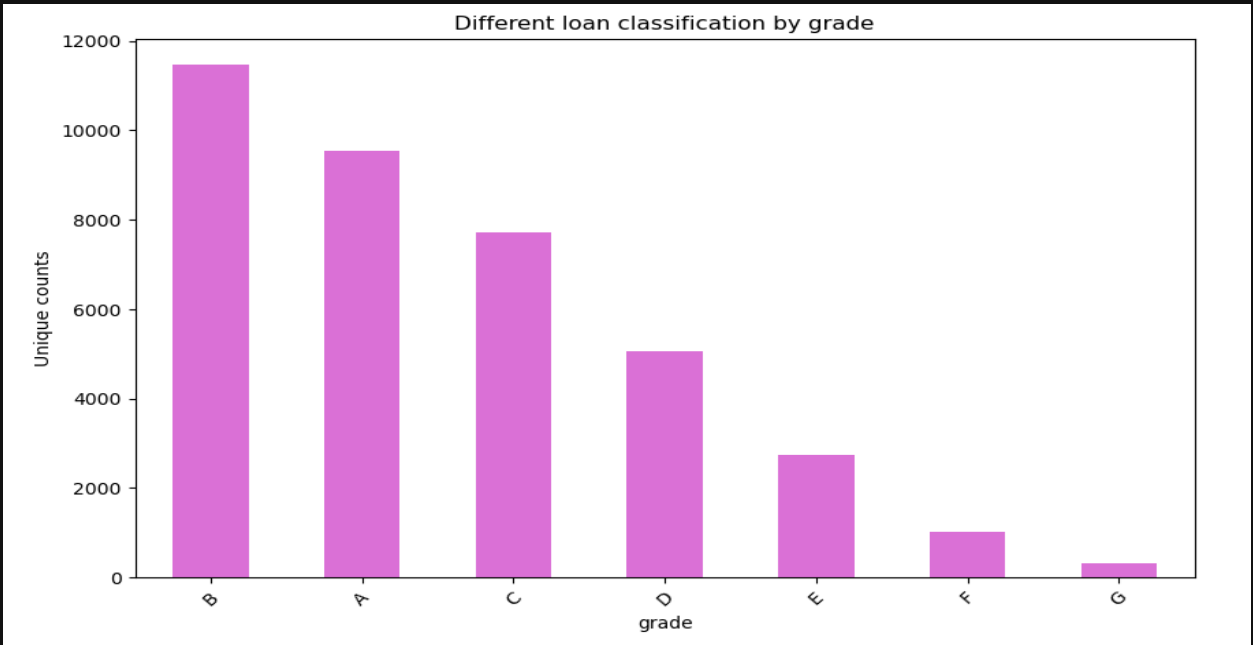
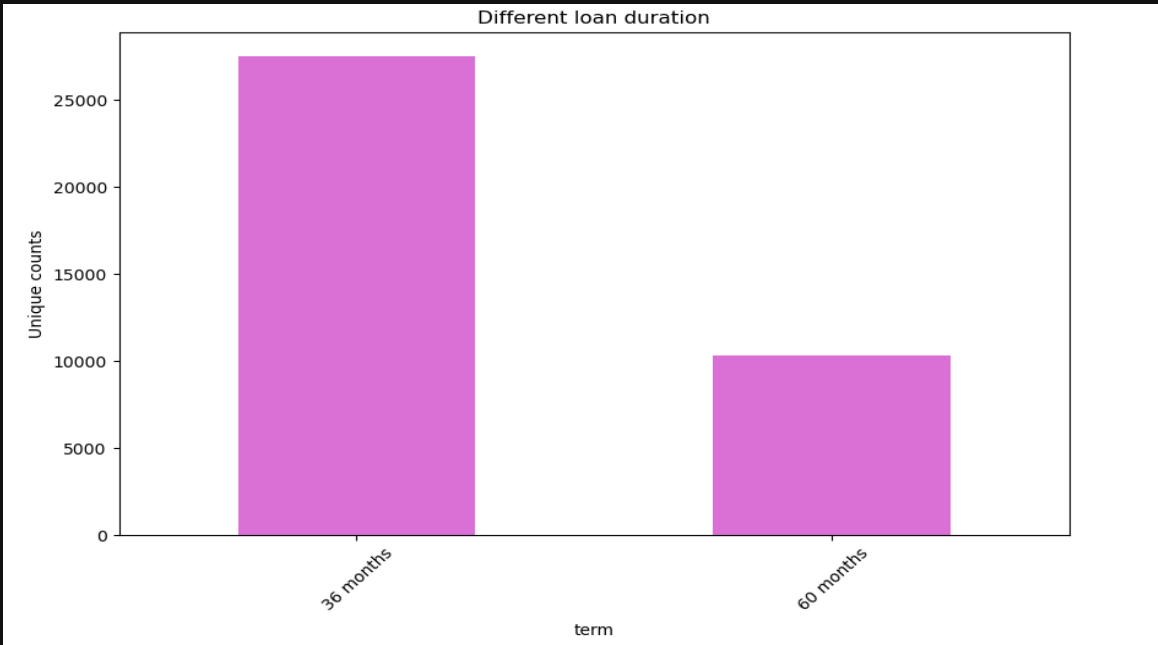
Distribution of total credit revolving balance

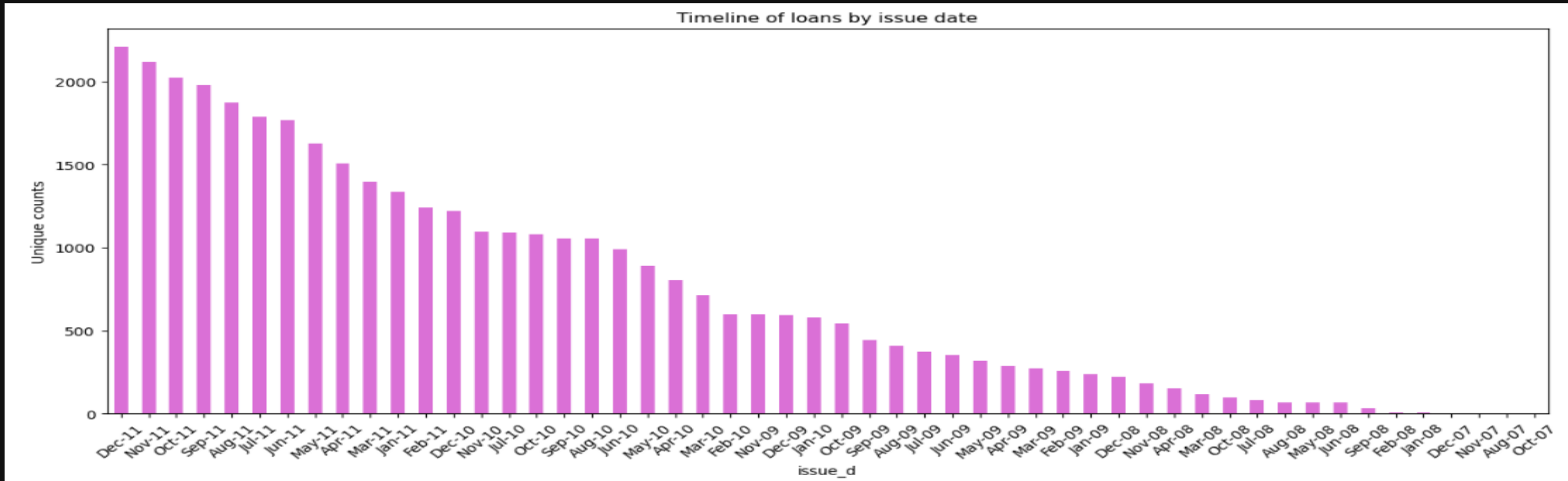
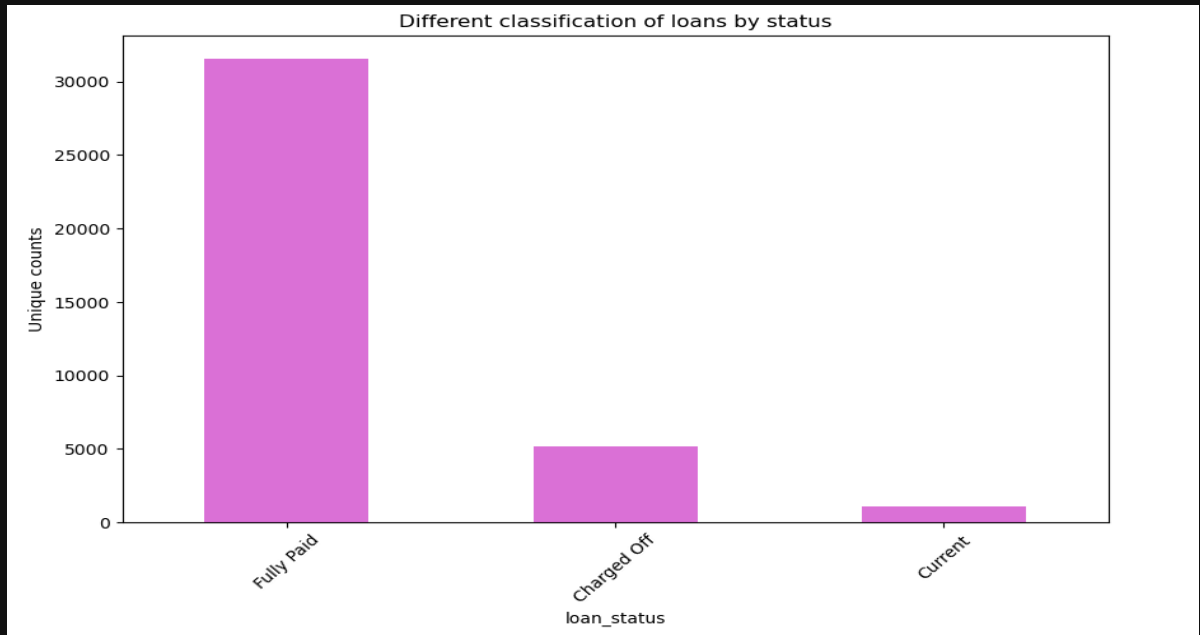
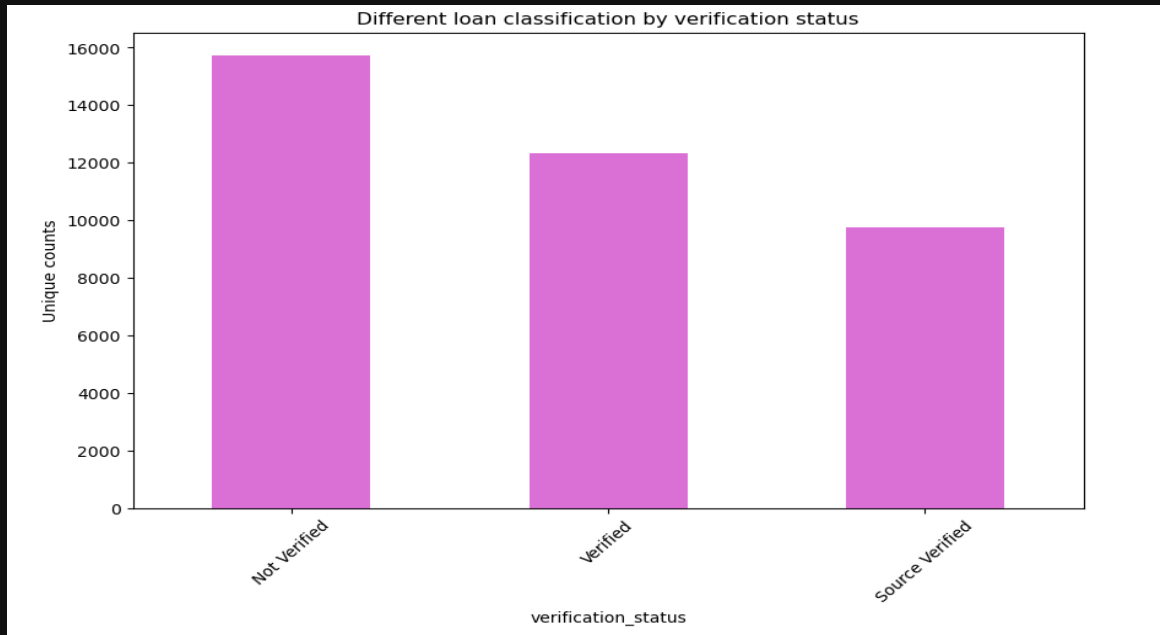


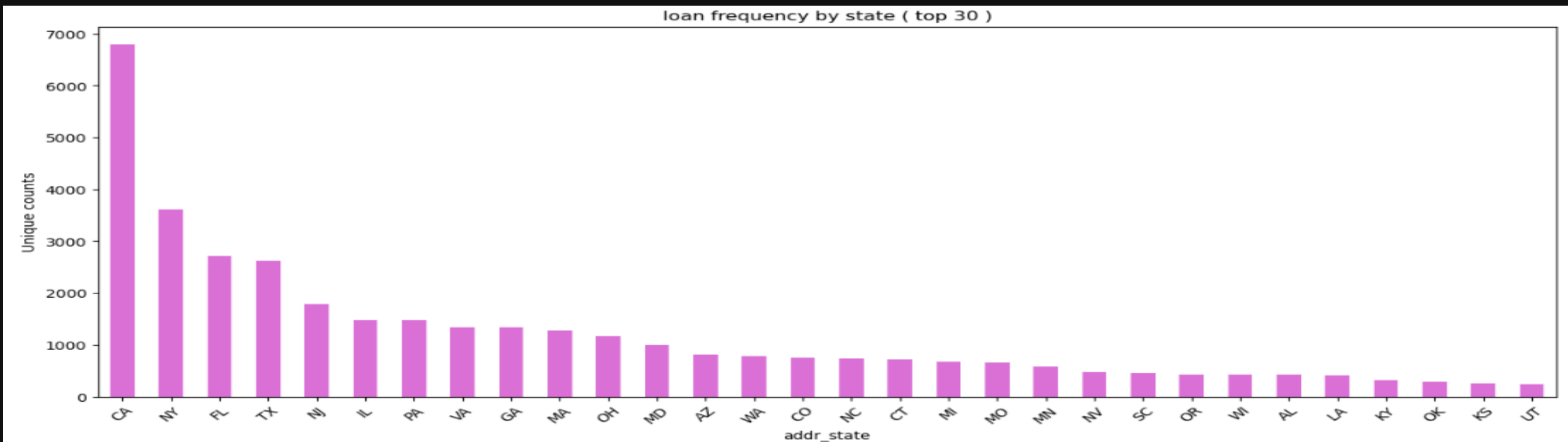
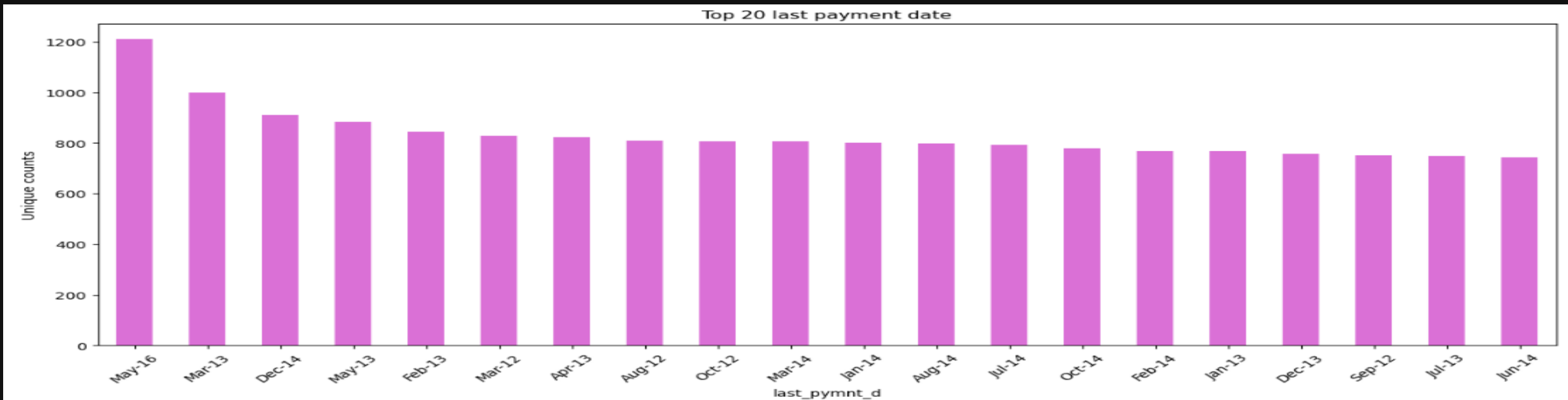
Distribution of revolving line utilization rate

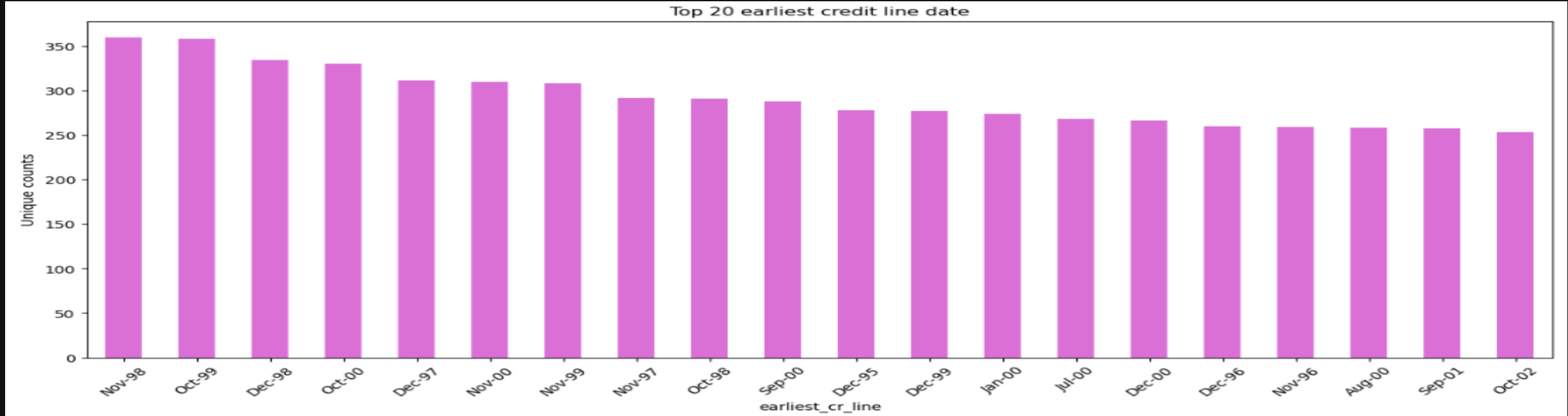
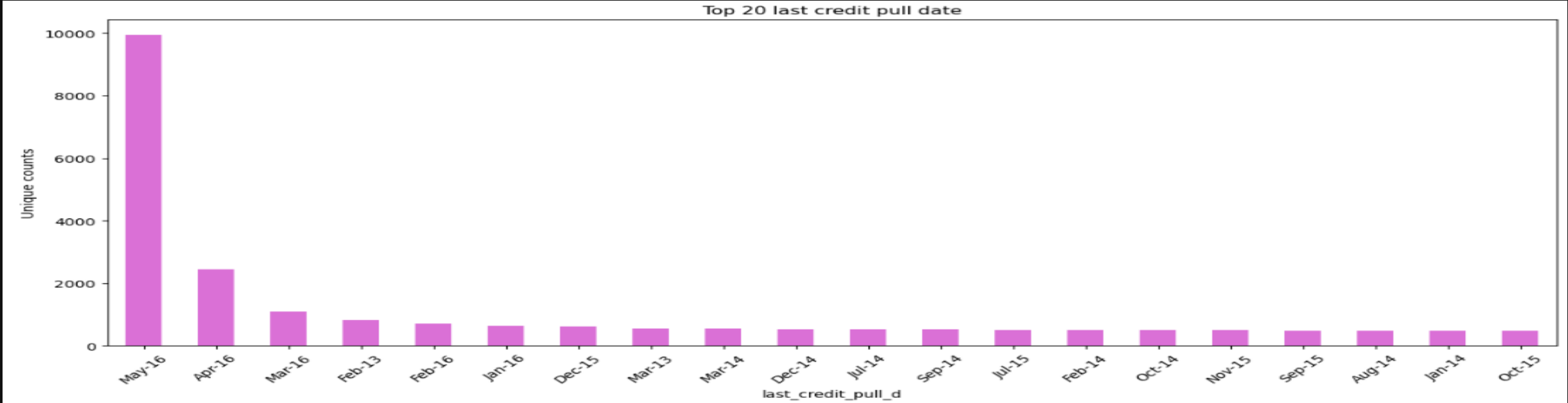


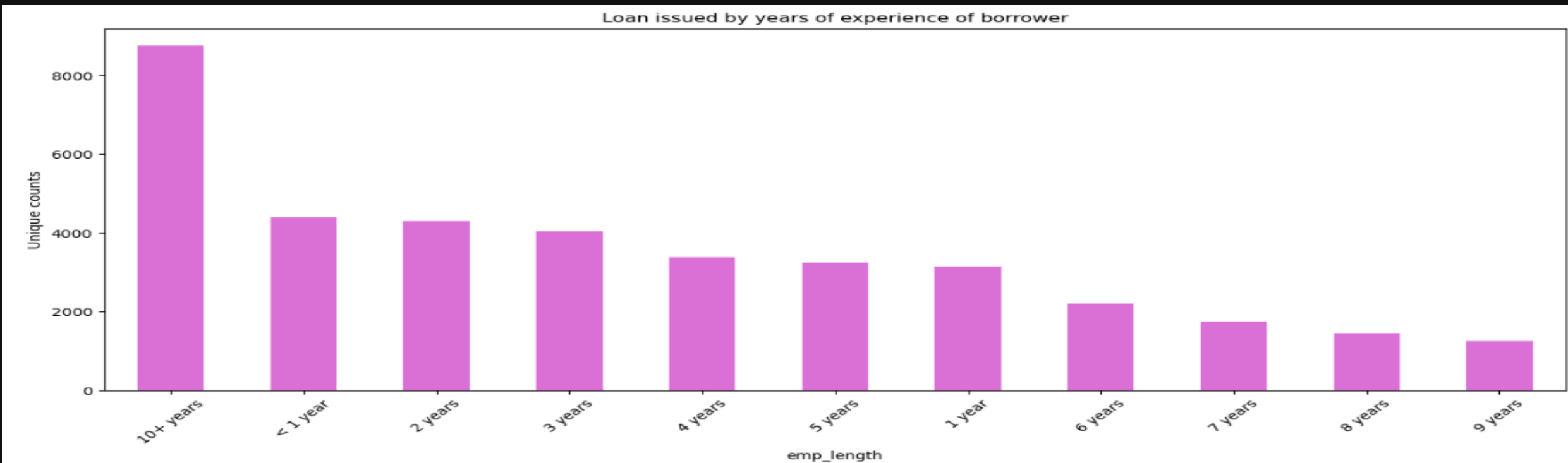
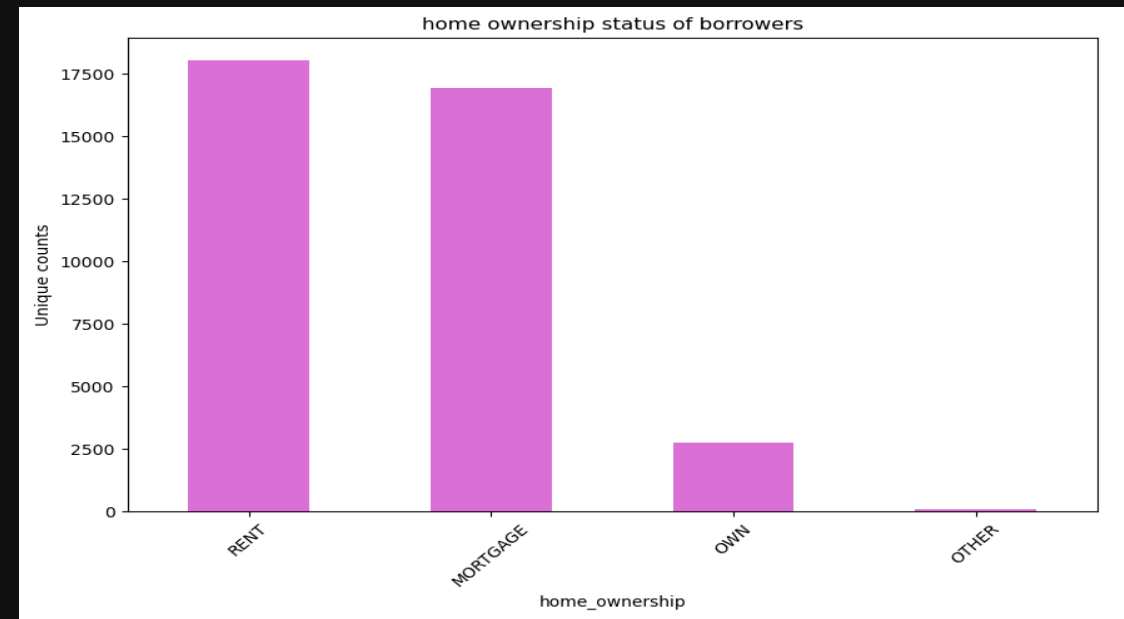
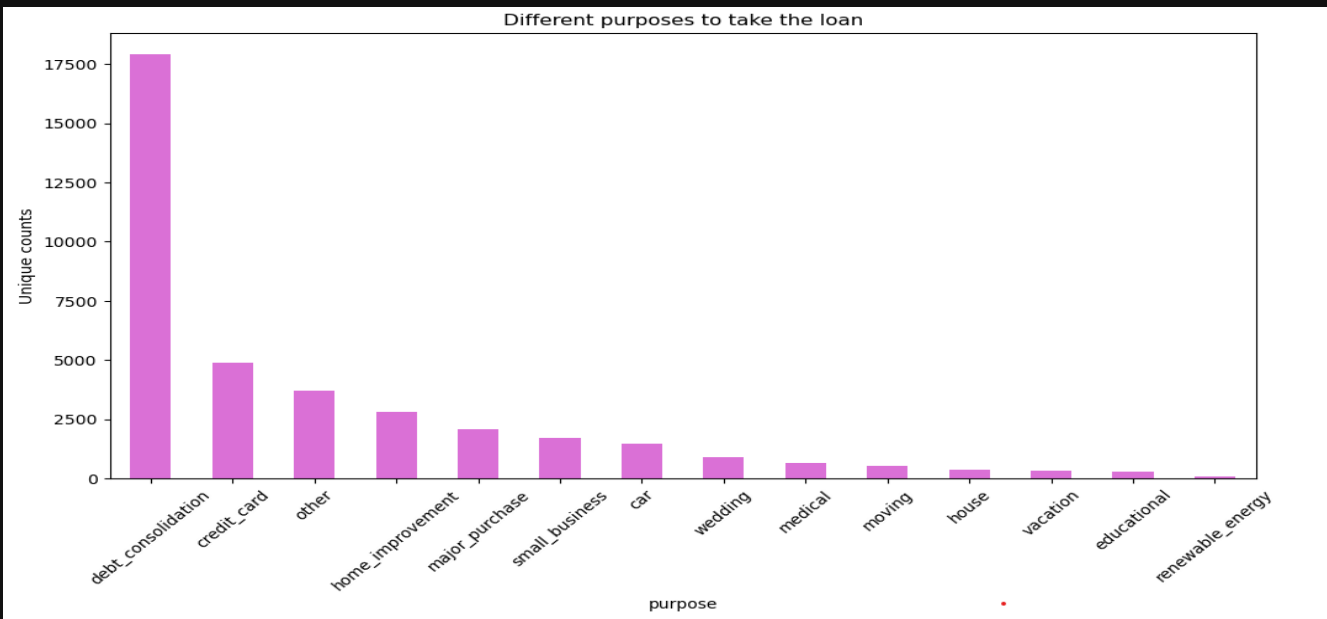
Categorical values







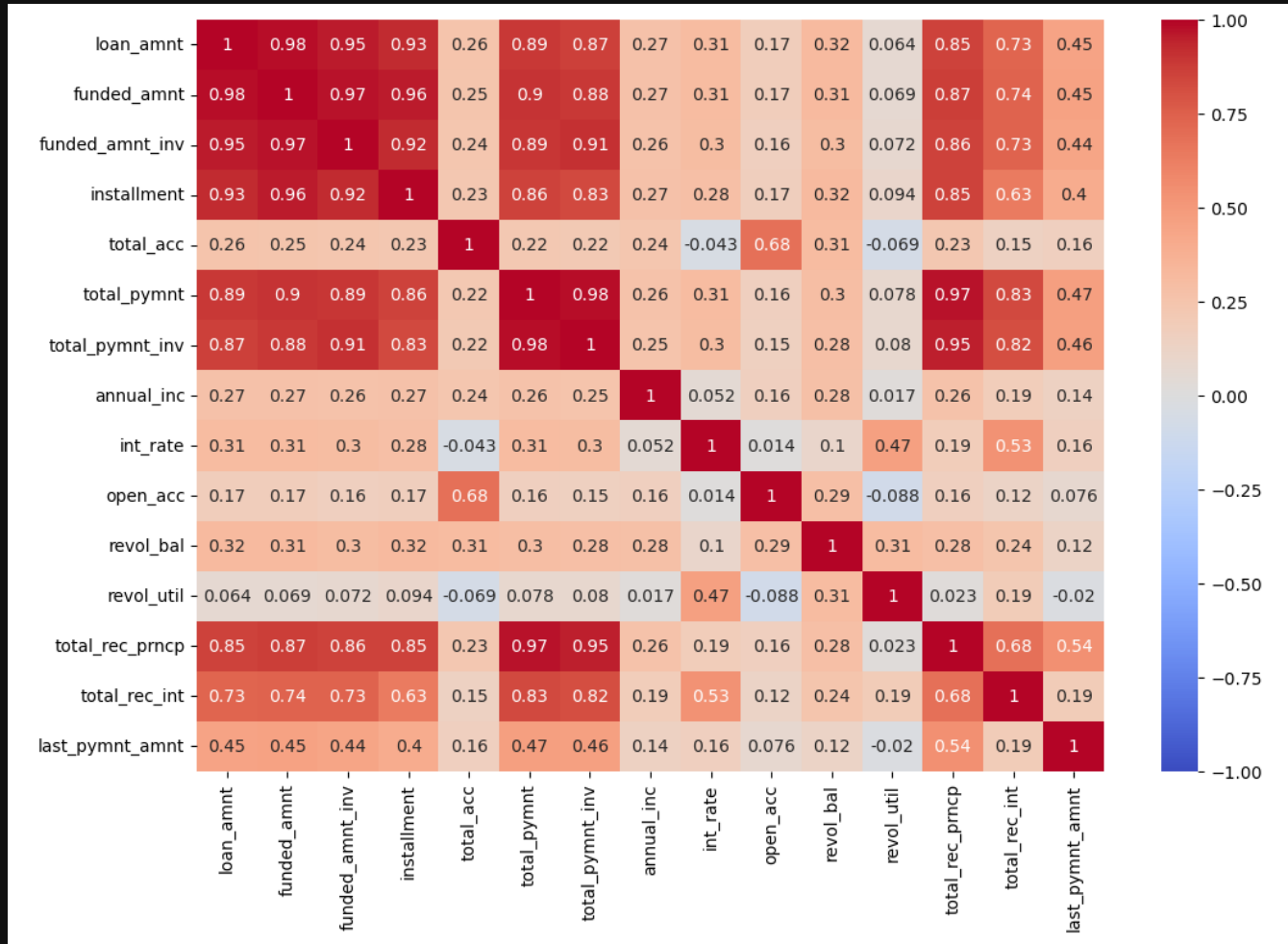




BIVARIATE ANALYSIS NUMERICAL VARIABLE

Relationship analysis – numerical variables

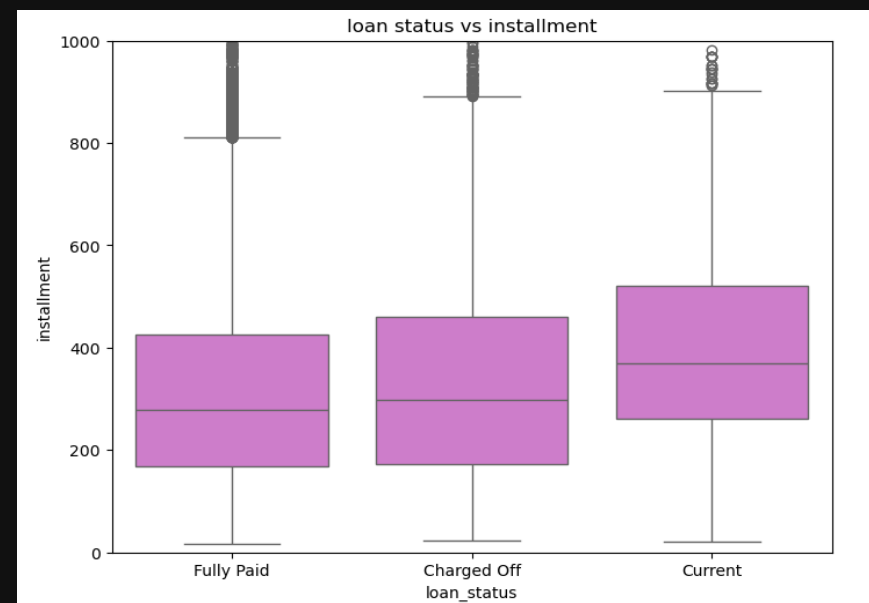
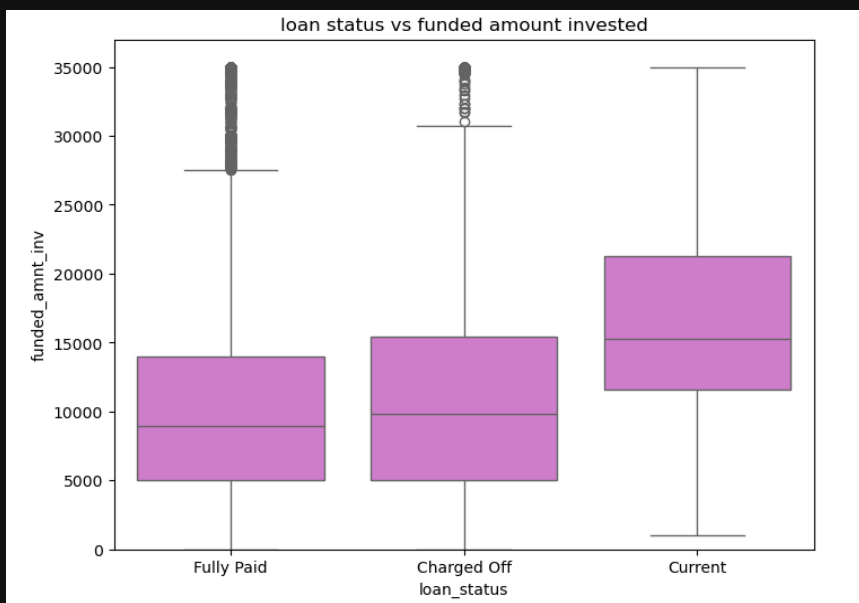
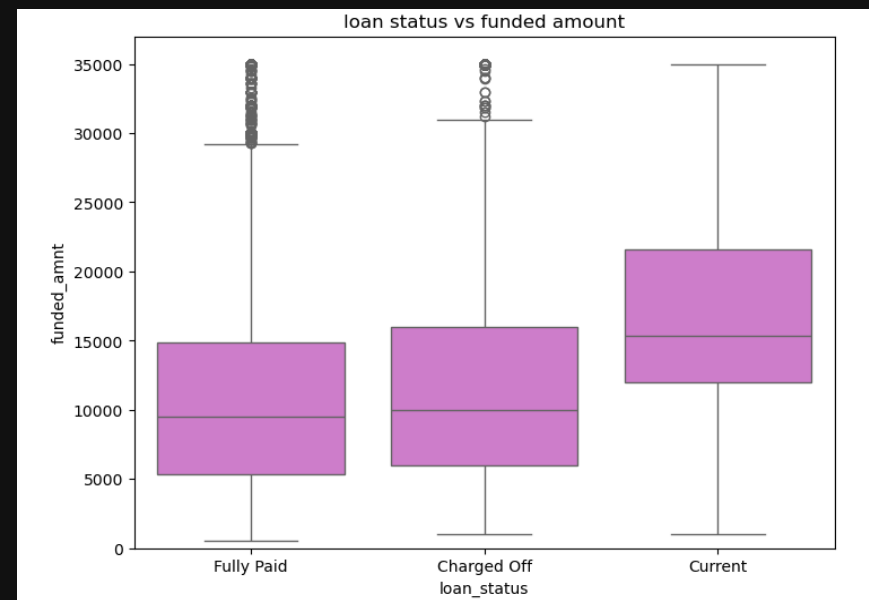
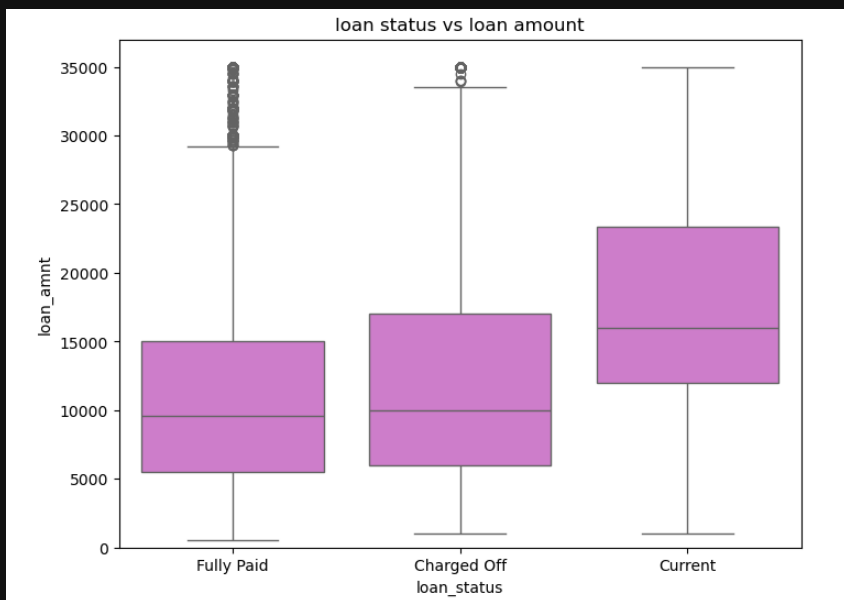
Correlation Matrix of important numerical variables

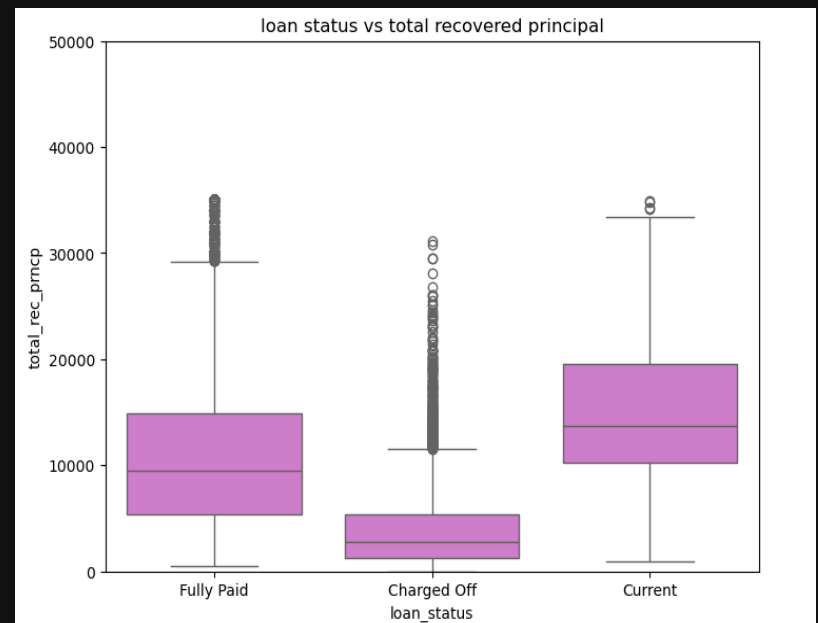
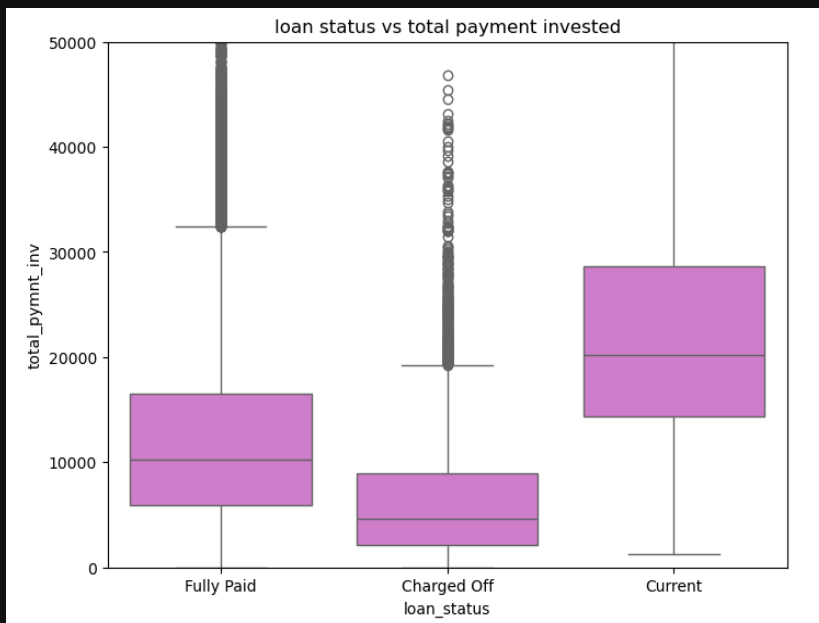
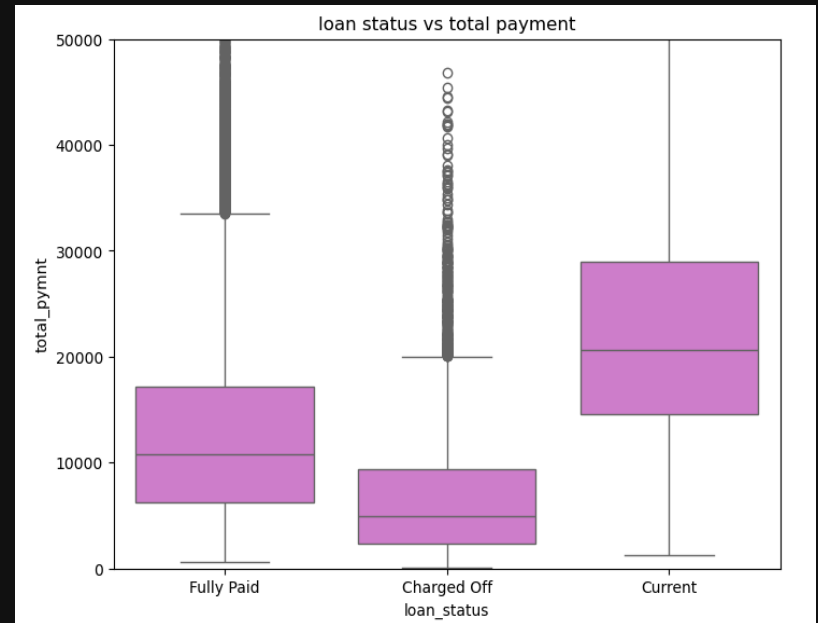
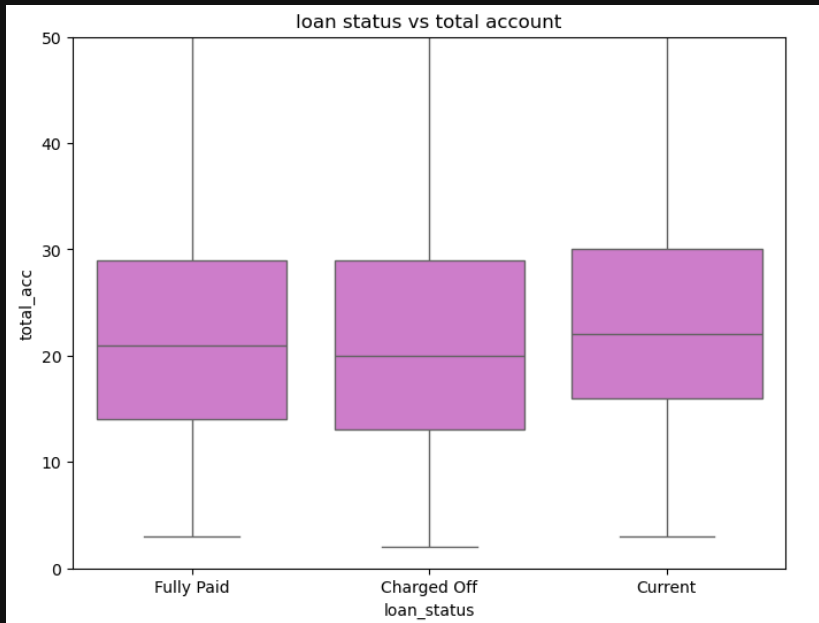


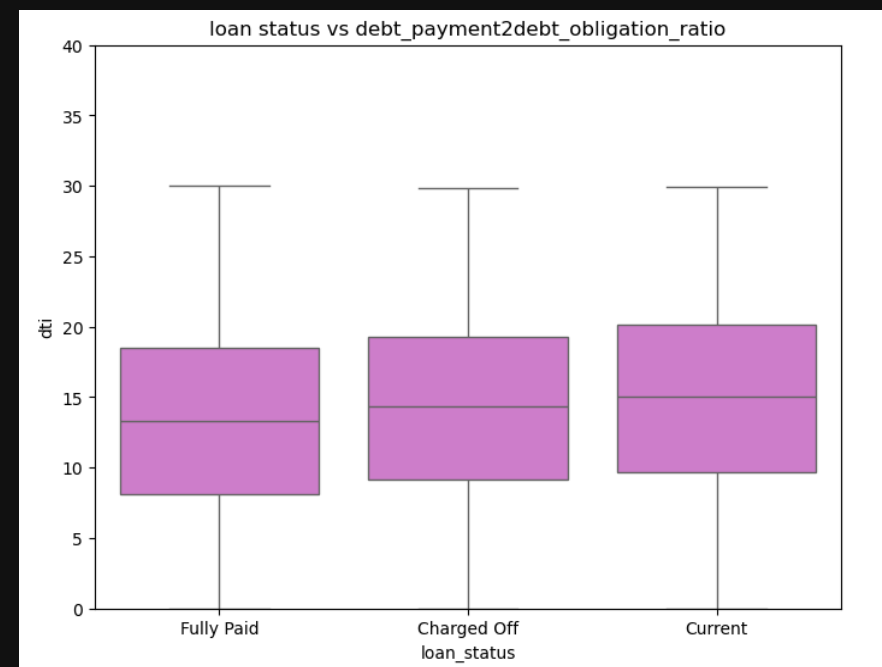
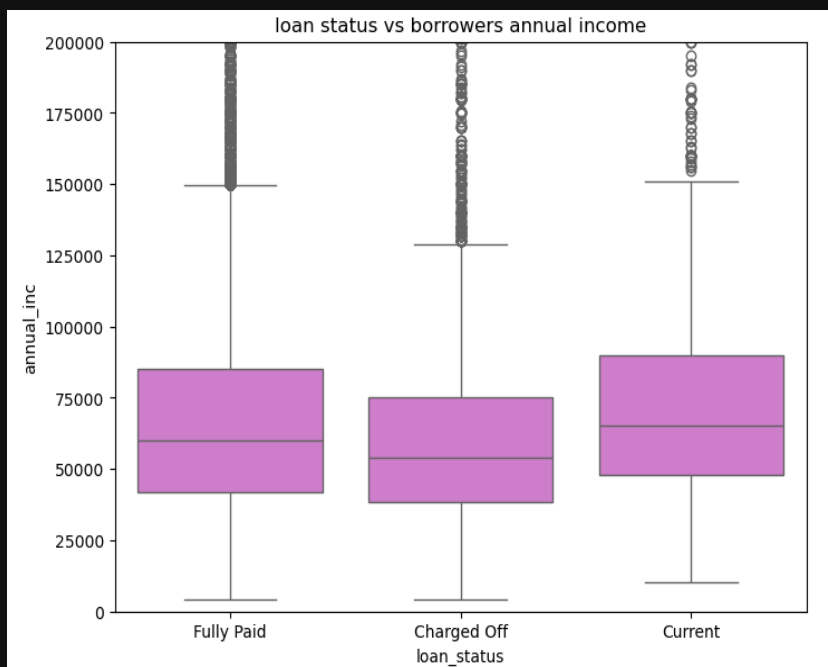
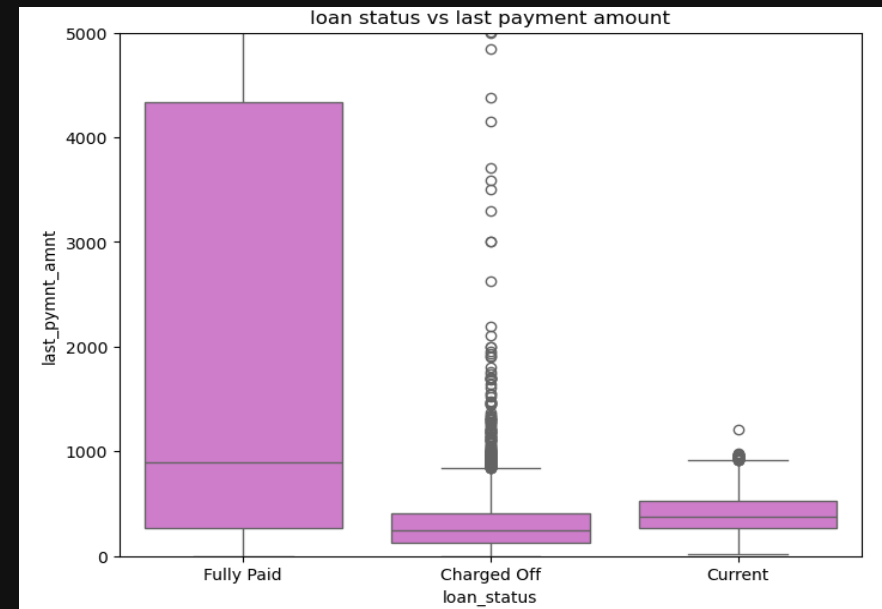
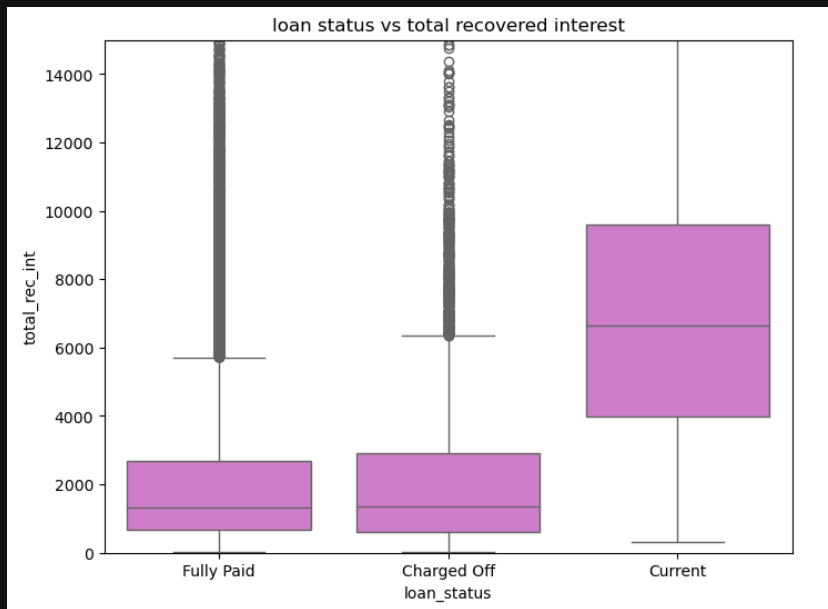
Inference :

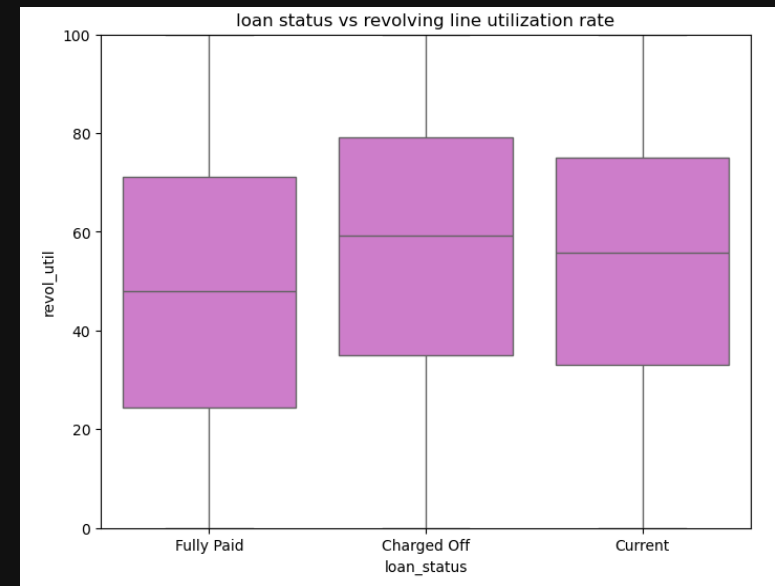
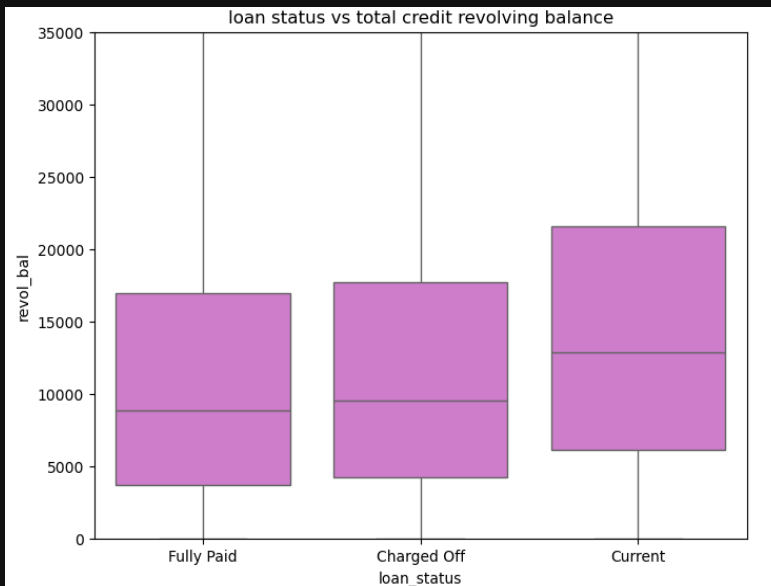
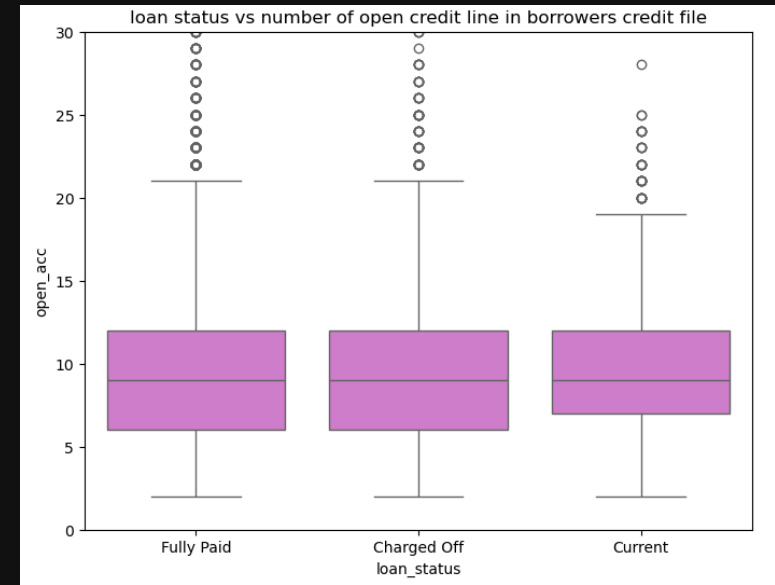
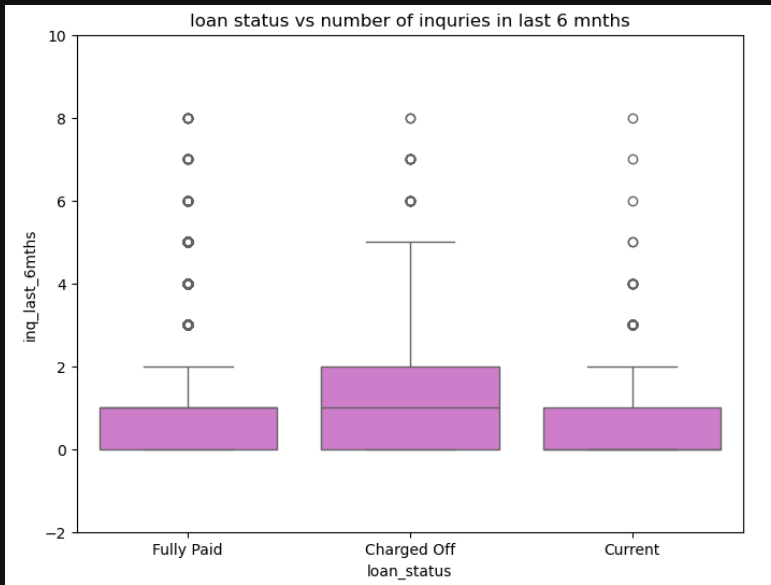
1. loan_amnt has very strong + correlation with funded_amnt , funded_amnt_inv which is an Indication that the CFC is operationally well run to meet its commitments.
2. loan_amnt has very strong + correlation with installment which is obvious since higher loan_amnt will lead to higher installment controlled for term
3. loan_amnt has strong + correlation with total_pymnt , total_pymnt_inv which is a strong indicator that the CFC is operationally efficient and prioritizes investors
4. loan_amnt has weak correlation with annual_inc which needs to be considered while giving loan
5. Loan_amnt has very strong + correlation with total_rec_prncp , total_rec_int , last_pymnt_amnt which is a great indicator that the CFC is great at collection management reducing structural risk
6. Total_pymnt , total_pymnt_inv , total_rec_prncp , total_rec_int has strong + correlation as they are leading and lagging indicator of efficient collection system
7. Int_rate has strong + correlation with revol_util which may be an indicator that the CFC is catering to a segment which is loan starved from other sources or their interest rate is still substantially higher
8. total_acc has strong + correlation with open_acc which is an indicator that most of the credit files are open , which is an indicator of strong loan book
9. Int_rate has strong + correlation with total_rec_int which is an obvious conclusion since higher interest rate leads to higher interest recovery considering its not a bad loan

Bivariate analysis :
Loan status with numerical variables



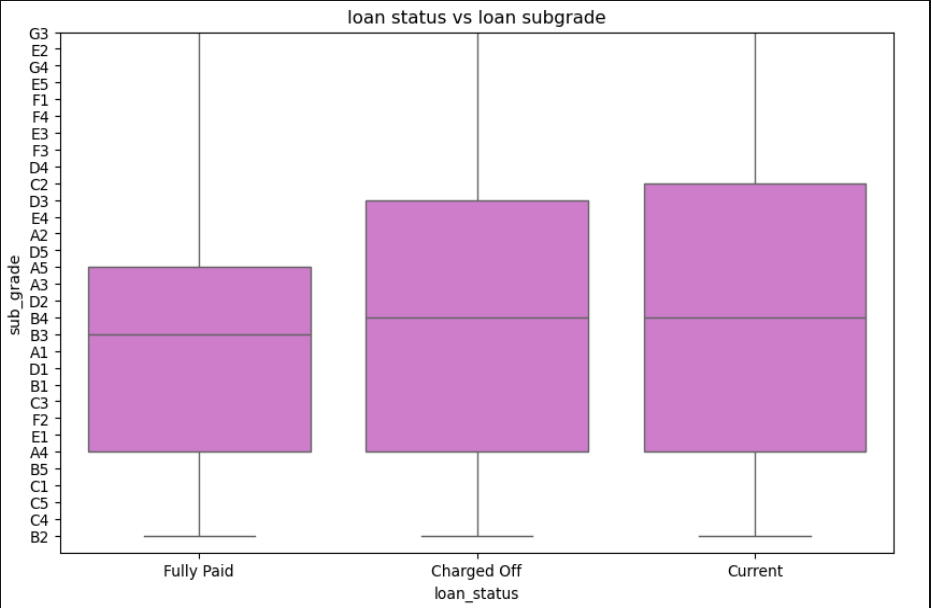
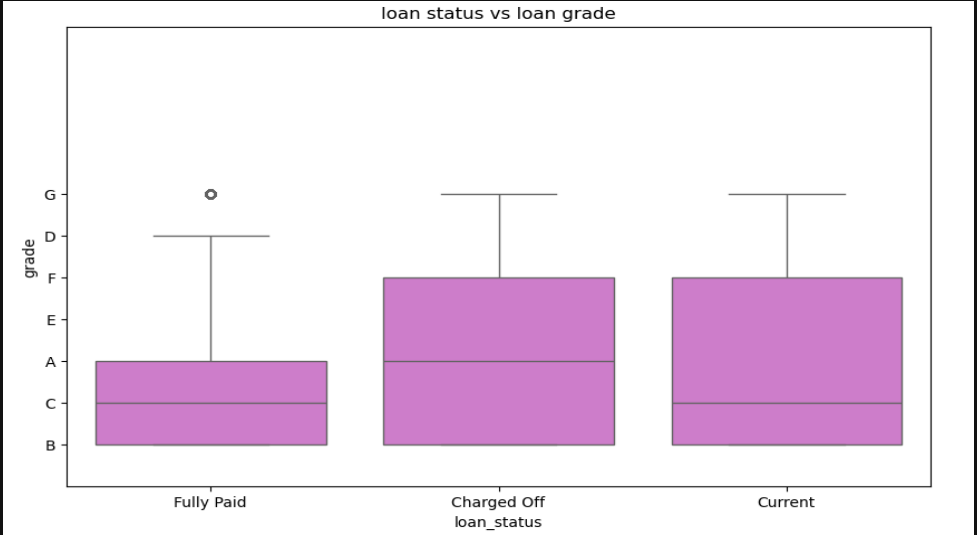
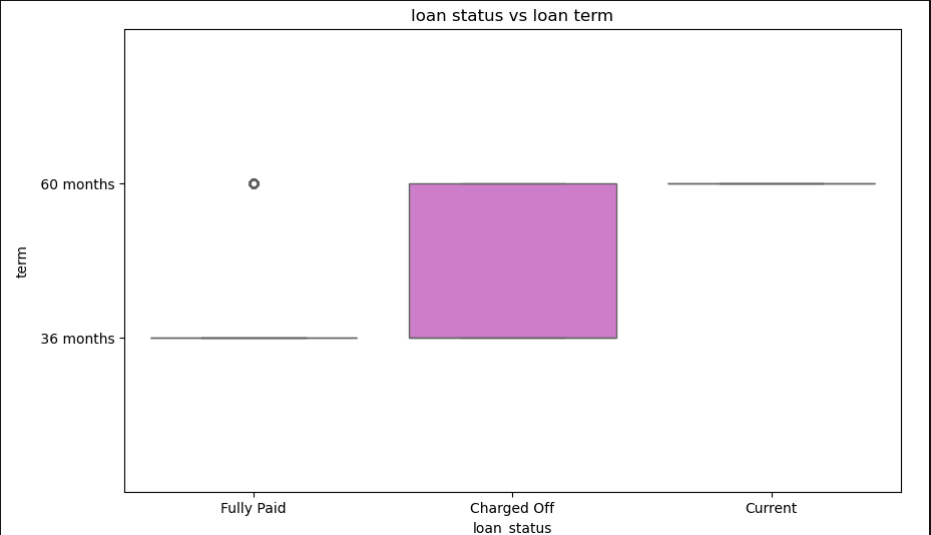


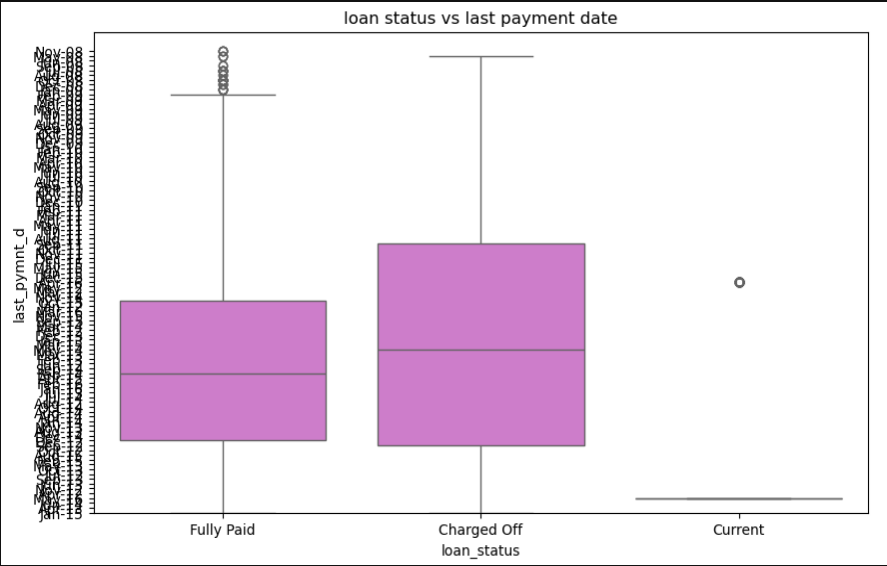
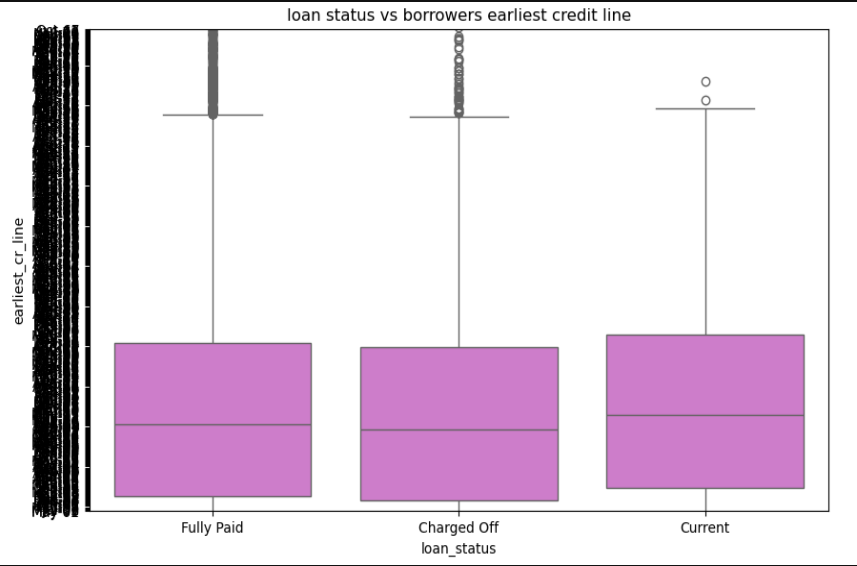
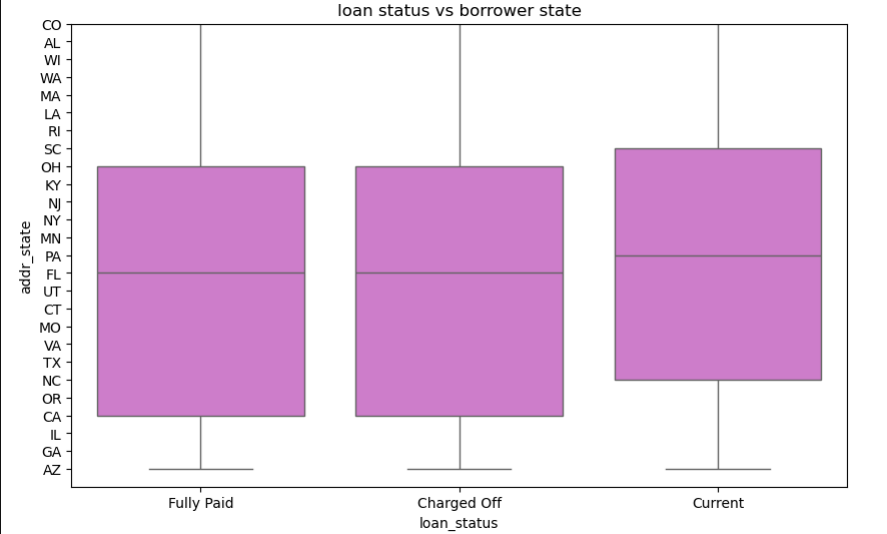
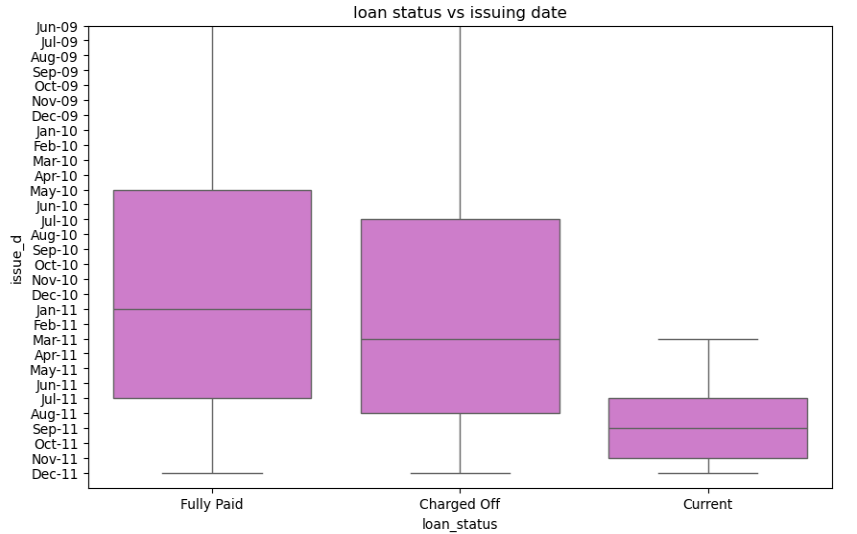


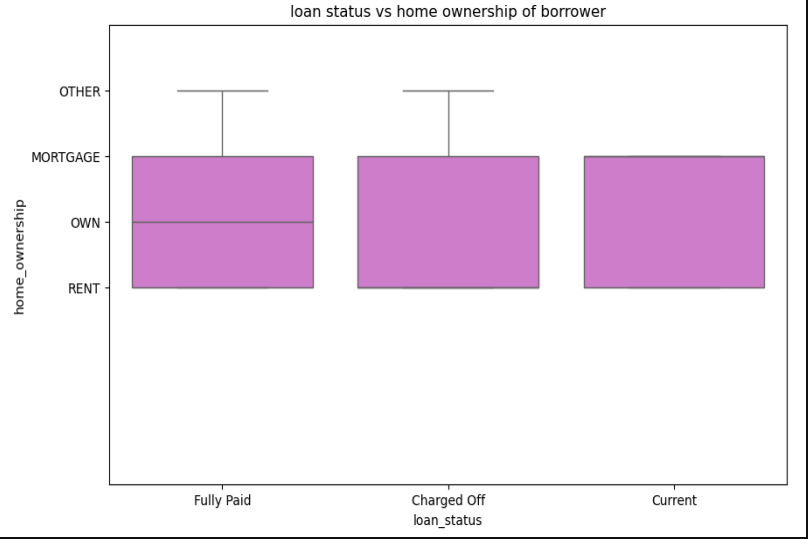
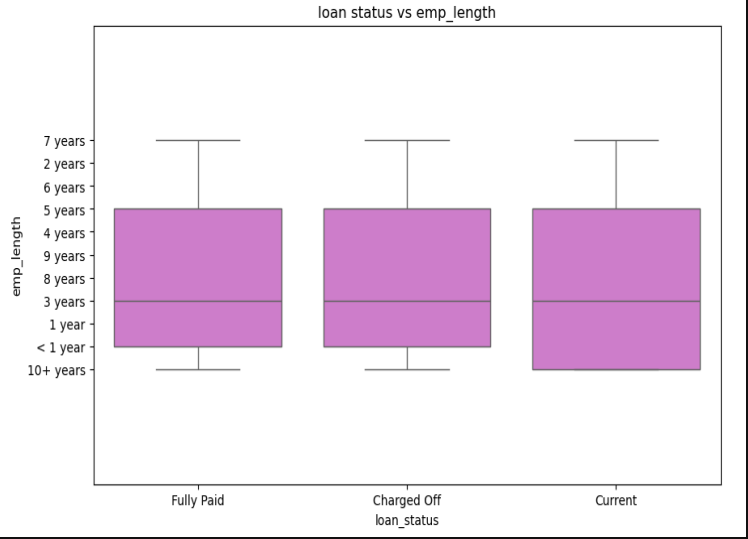
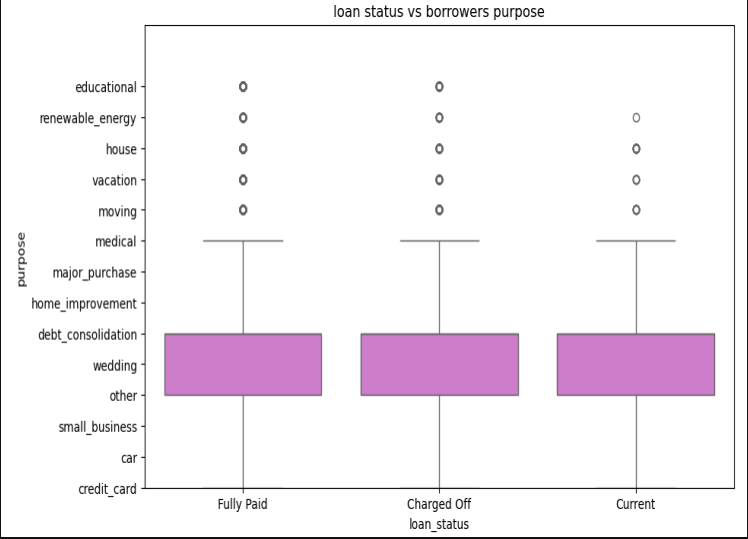
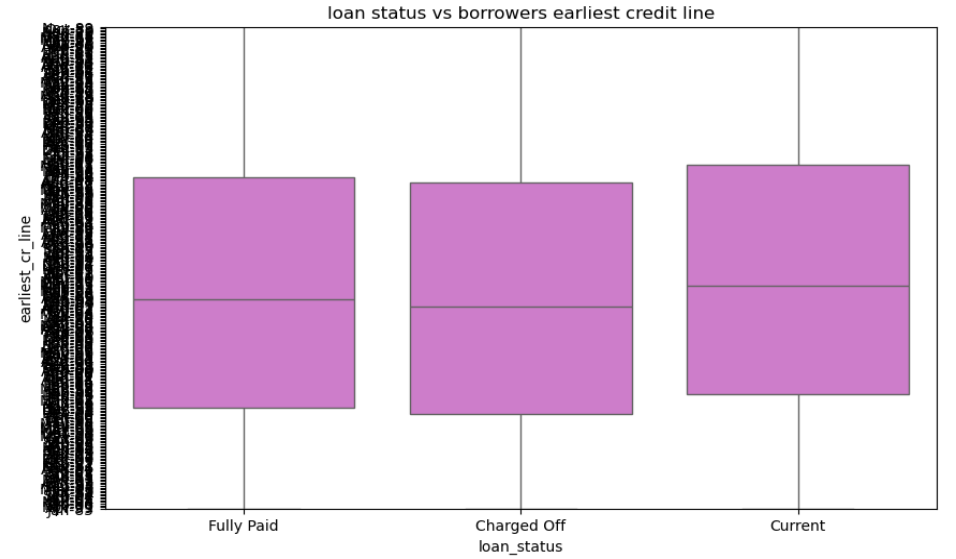
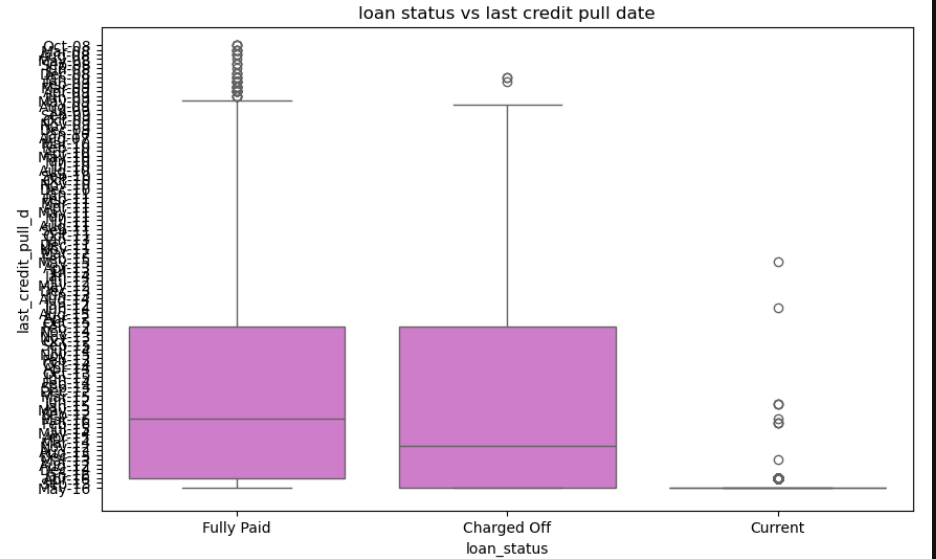


BIVARIATE ANALYSIS CATEGORICAL VARIABLE

Bivariate analysis :
Loan status with Categorical variables

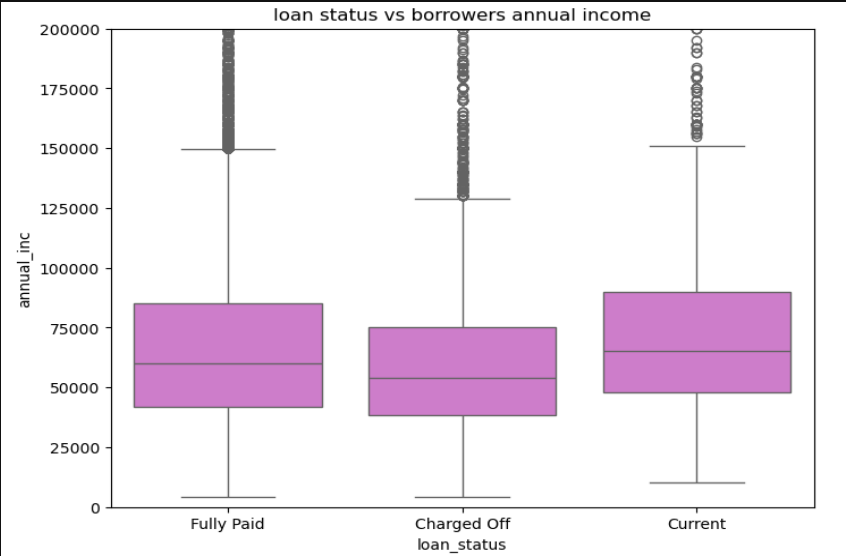
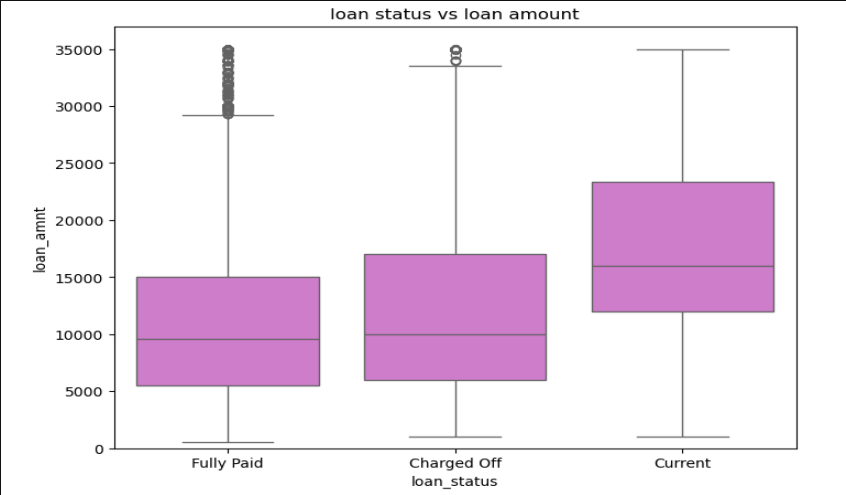






Final Insight

INSIGHT 1



	count	mean
loan_status		
Charged Off	5203.0	12296.290602
Current	1098.0	17182.604736
Fully Paid	31534.0	10964.766601

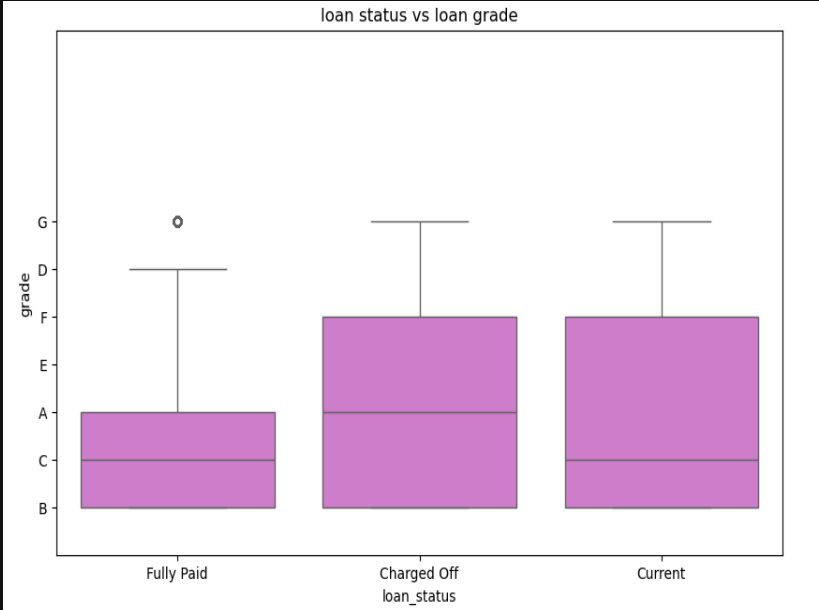
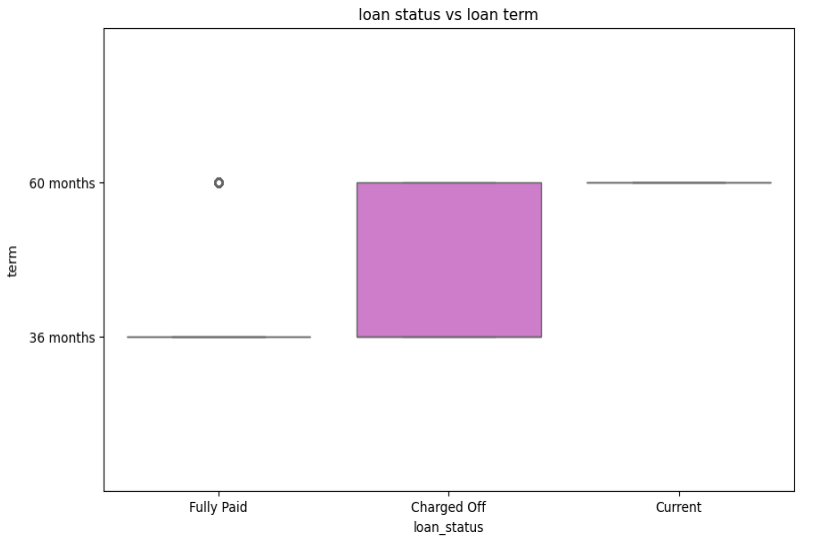
Inference :

1. Borrowers on average who take smaller amount loan generally have higher probability of “Fully Paid”
2. Borrowers on average who has higher annual income generally shows higher probability of “Fully Paid”

Action :

Give higher weightage to annual income of a borrower while providing loan and consider smaller size loan amount products

INSIGHT 2



	loan_status	term	count
0	Charged Off	36 months	2895
1	Charged Off	60 months	2308
2	Current	60 months	1098
3	Fully Paid	36 months	24603
4	Fully Paid	60 months	6931

Inference :

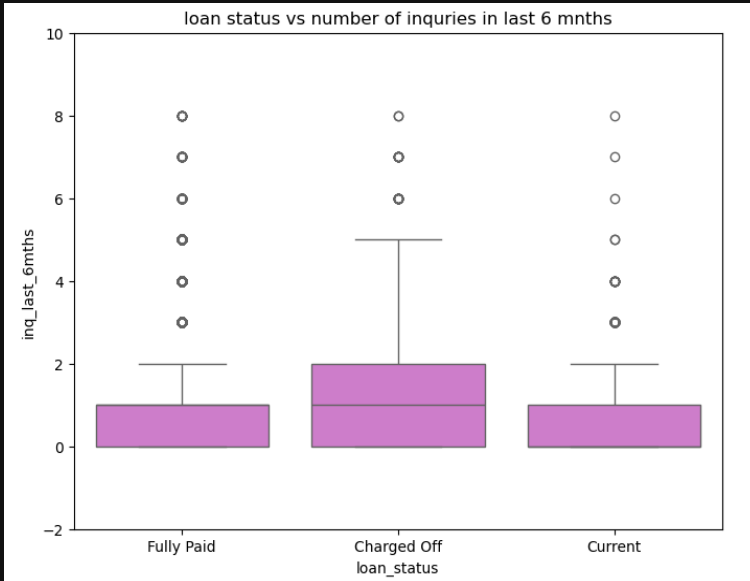
1. Borrowers on average who take loans for smaller duration generally have higher probability of “Fully Paid”
2. Borrowers on average who take Grade A , B , C , D has higher probability of “Fully Paid”

	loan_status	grade	count
0	Charged Off	A	546
1	Charged Off	B	1313
2	Charged Off	C	1229
3	Charged Off	D	1036
4	Charged Off	E	679
5	Charged Off	F	304
6	Charged Off	G	96
7	Current	A	35
8	Current	B	327
9	Current	C	257
10	Current	D	216
11	Current	E	175
12	Current	F	71
13	Current	G	17
14	Fully Paid	A	8966
15	Fully Paid	B	9823
16	Fully Paid	C	6223
17	Fully Paid	D	3803
18	Fully Paid	E	1888
19	Fully Paid	F	635
20	Fully Paid	G	196

Action :

Try to give more small term loans of Grade : A , B , C , D and drop grade (E , F , G) loans or charge higher interest rate to reduce structural risk

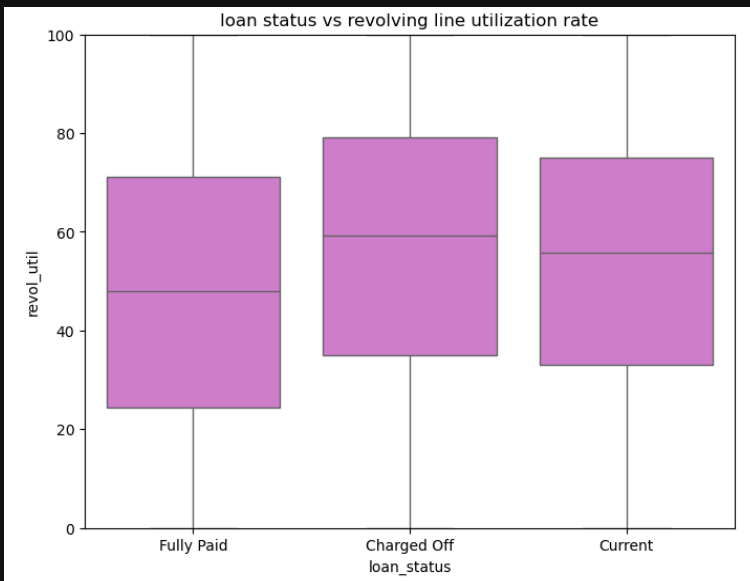
INSIGHT 3



	count	mean
loan_status		
Charged Off	5203.0	1.048241
Current	1098.0	0.793260
Fully Paid	31534.0	0.839443

Inference :

1. Borrowers on average who inquired least in last 6 month have higher probability of “Fully Paid” , the reverse can act as a good leading indicator for collection team to work on
2. Borrowers on average who has lower utilization of approved loan amount generally shows higher probability of “Fully Paid”

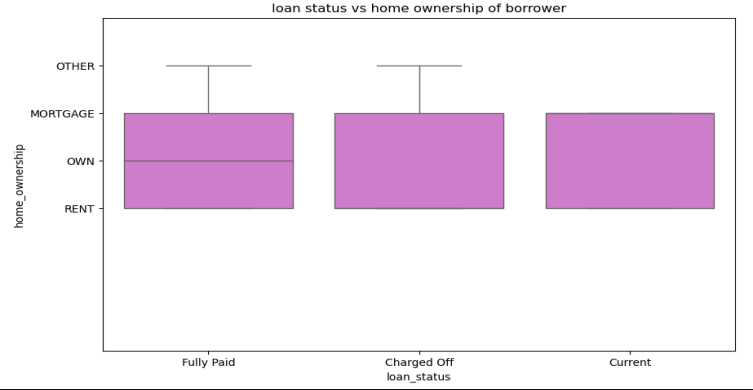
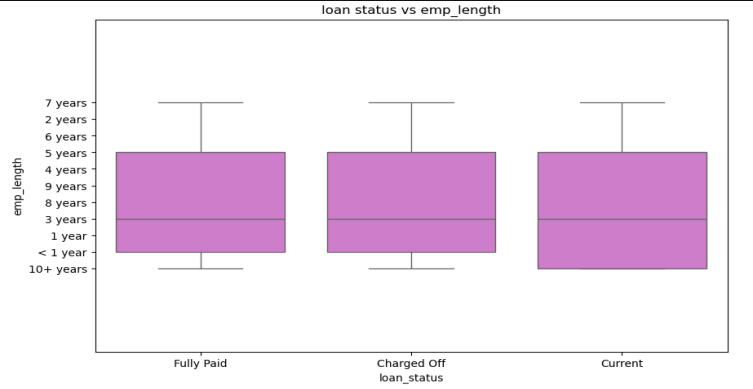
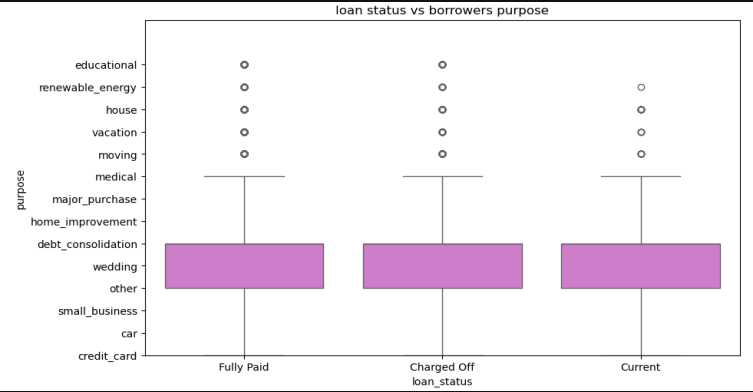


	count	mean
loan_status		
Charged Off	5203.0	55.990680
Current	1098.0	53.683160
Fully Paid	31534.0	47.837262

Action :

1. Digitize the complete customer support / helpdesk functionality to ensure we have a centralized view of all historical enquiries of a customer / lead .
2. Continuously monitor the current utilization rate of different customers and based on threshold value flag it as high risk for collection team to work on

INSIGHT 4



loan_status	purpose	count
0 Charged Off	credit_card	151
1 Charged Off	debt_consolidation	455
2 Charged Off	educational	27
3 Charged Off	home_improvement	316
4 Charged Off	house	208
5 Charged Off	major_purchase	201
6 Charged Off	medical	99
7 Charged Off	moving	61
8 Charged Off	other	571
9 Charged Off	renewable_energy	17
10 Charged Off	small_business	436
11 Charged Off	wedding	50
12 Charged Off	wedding	63
13 Charged Off	wedding	49
14 Current	credit_card	101
15 Current	debt_consolidation	507
16 Current	home_improvement	94
17 Current	house	14
18 Current	major_purchase	26
19 Current	medical	12
20 Current	moving	7
21 Current	other	120
22 Current	renewable_energy	1
23 Current	small_business	72
24 Current	wedding	4
25 Current	wedding	31
26 Fully Paid	credit_card	1276
27 Fully Paid	debt_consolidation	4298
28 Fully Paid	educational	14750
29 Fully Paid	home_improvement	2116
30 Fully Paid	house	201
31 Fully Paid	major_purchase	1666
32 Fully Paid	medical	62
33 Fully Paid	moving	460
34 Fully Paid	other	2017
35 Fully Paid	renewable_energy	76
36 Fully Paid	small_business	1212
37 Fully Paid	wedding	321
38 Fully Paid	wedding	806

loan_status	emp_length	count
0 Charged Off	< 1 year	430
1 Charged Off	10+ years	1203
2 Charged Off	2 years	549
3 Charged Off	3 years	542
4 Charged Off	4 years	441
5 Charged Off	5 years	443
6 Charged Off	6 years	298
7 Charged Off	7 years	257
8 Charged Off	8 years	106
9 Charged Off	9 years	153
10 Charged Off	< 1 year	601
11 Current	< 1 year	71
12 Current	10+ years	201
13 Current	2 years	97
14 Current	3 years	83
15 Current	4 years	64
16 Current	5 years	88
17 Current	6 years	61
18 Current	7 years	62
19 Current	8 years	44
20 Current	9 years	32
21 Current	< 1 year	75
22 Fully Paid	< 1 year	2652
23 Fully Paid	10+ years	7066
24 Fully Paid	2 years	2647
25 Fully Paid	3 years	3366
26 Fully Paid	4 years	2842
27 Fully Paid	5 years	2704
28 Fully Paid	6 years	1934
29 Fully Paid	7 years	1428
30 Fully Paid	8 years	1209
31 Fully Paid	9 years	1053
32 Fully Paid	< 1 year	2721

loan_status	home_ownership	count
0 Charged Off	MORTGAGE	2183
1 Charged Off	OTHER	18
2 Charged Off	OWN	382
3 Charged Off	RENT	2620
4 Current	MORTGAGE	617
5 Current	OWN	71
6 Current	RENT	410
7 Fully Paid	MORTGAGE	14138
8 Fully Paid	OTHER	78
9 Fully Paid	OWN	2318
10 Fully Paid	RENT	15000

Inference :

1. Most of our loans are taken for debt consolidation , wedding and other categories
2. Most of our customers experience is extreme either closer to 1 year or over 10 years
3. Most of our customers have home mortgage or at Rent

Action :

1. A loan product can be created for high ticket productive asset which will structurally have low risk

OVERALL INSIGHT SUMMARY



Giving small amount of loan to high income individual is very low risk



Giving short term loan of Grade A , B , C is very low on risk



Closely monitor loan utilization rate and last 6 month enquiries of a customer / lead , they are a good leading indicator of high risk loans



Product development opportunity exist to build a high ticket loan product for high experience customer , who doesn't own home

THANK YOU

APPENDIX