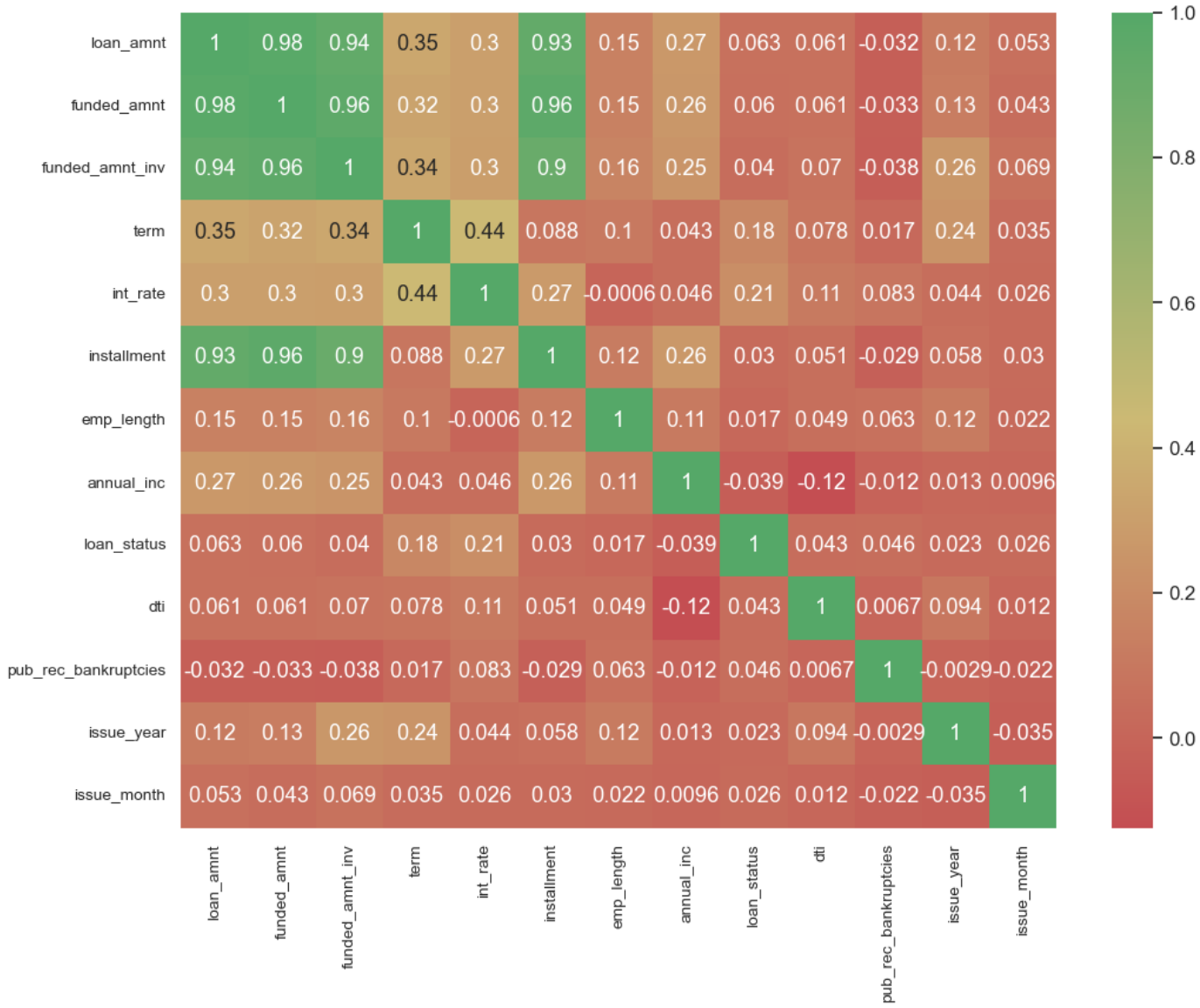
Data understanding and Clean up

- Totally 39717 Rows and 111 Columns
- Out of which
  - 56 columns have more than 90 percent missing values
  - 9 columns have the same value for the all rows
  - 22 columns are irrelevant
  - 7 columns are miscellaneous
  - Totally 94 columns are removed
- Final number of columns for analysis : 17
- Almost 85 % of the columns are filtered
- Row clean up : ‘Current’ Loan Status

Correlation Matrix

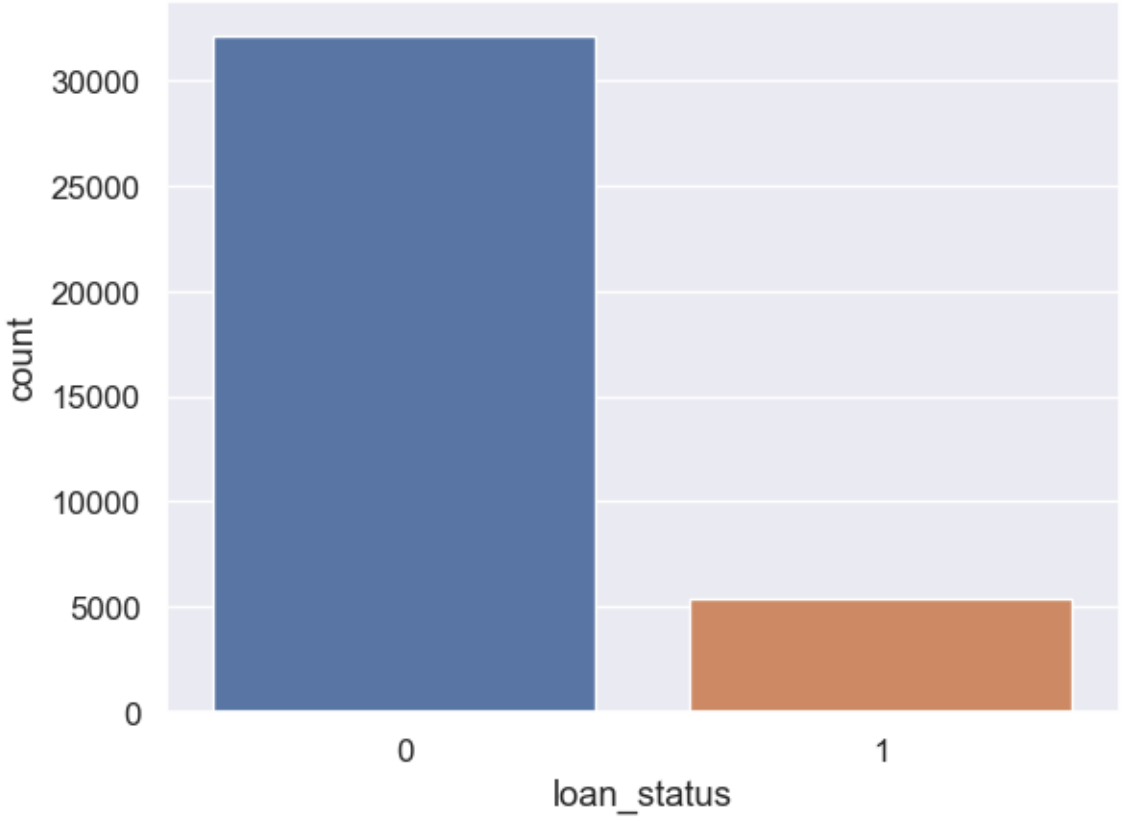
- Observations
  - Let us take sample variable as Loan status
  - The variables which are relatively higher correlation with respect to loan_status
    - Int_rate
    - term



Loan status

- The loan status ‘current’ is filtered
- 'Fully Paid’ is represented as 0
- ‘Charged off’ is represented as 1

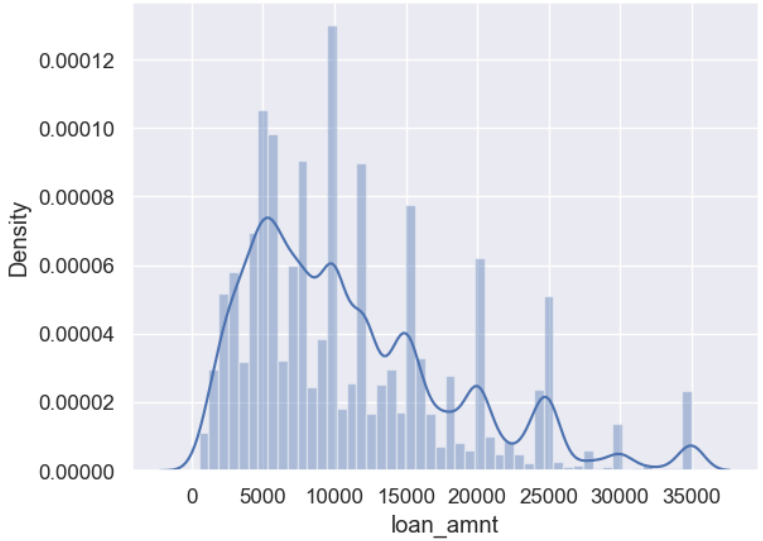
	Rows	Percentage after filtering
Fully Paid	32145	85.62
Charged Off	5399	14.38

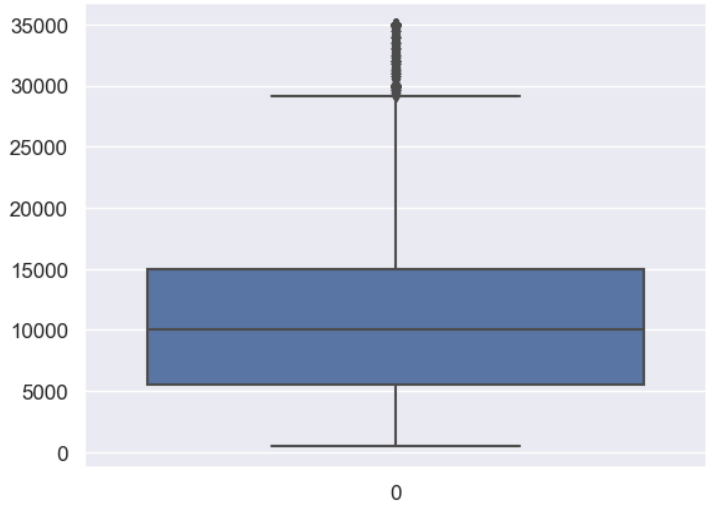


Fully Paid : Charged Off is almost 6 : 1

Loan Amount

count	37544.000000
mean	11119.329986
std	7354.098954
min	500.000000
25%	5500.000000
50%	10000.000000
75%	15000.000000
max	35000.000000

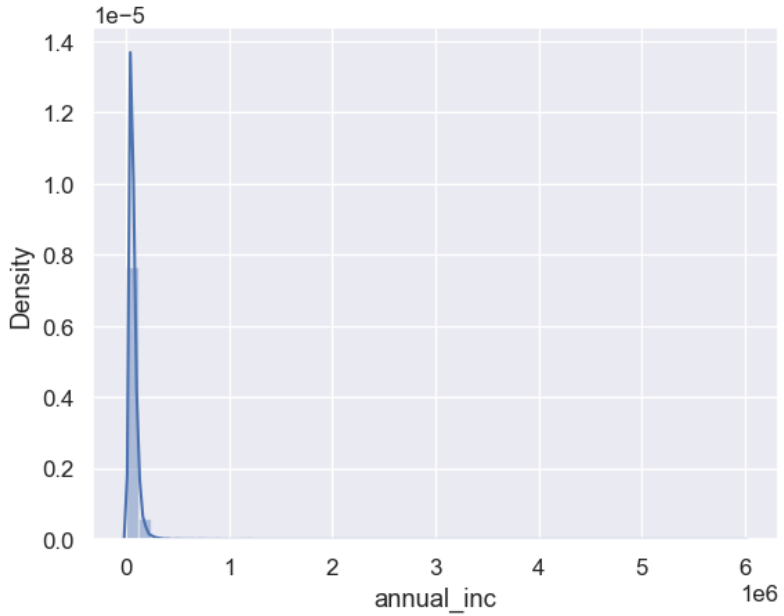




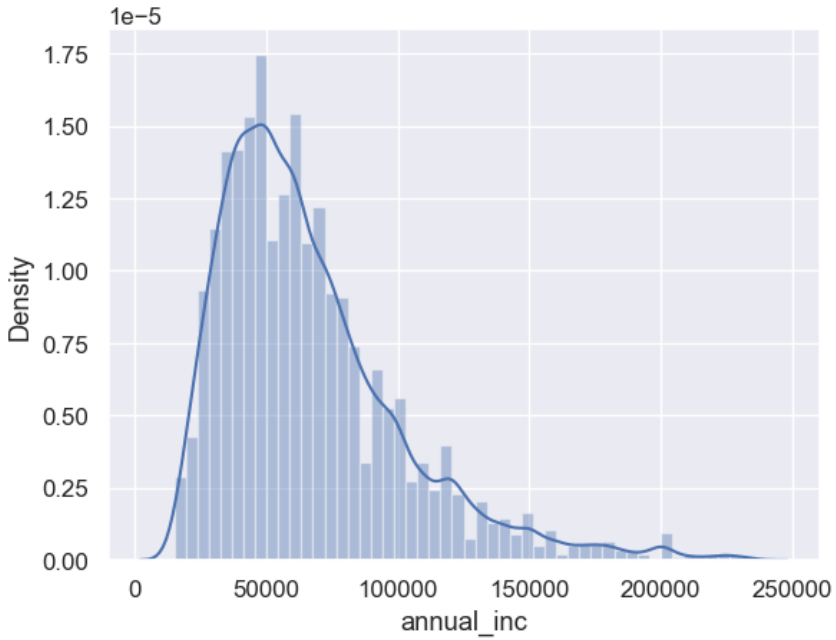
Observations

#1. Average Loan amount is 11119
#2. Most of the loan taken are in the range 5500 - 15000
#3. Maximum loan taken is 35000 and minimum is 500

Annual Income



After removal of outliers



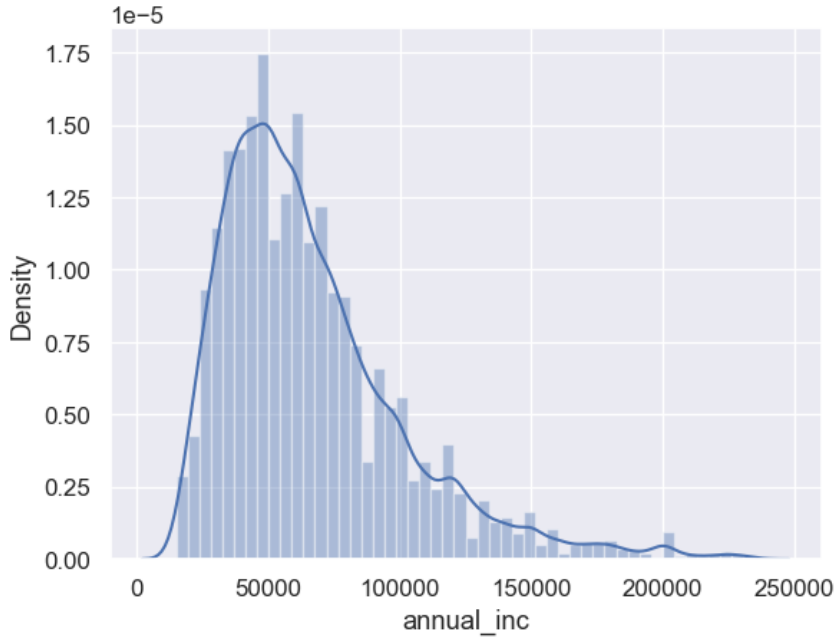
Annual Income

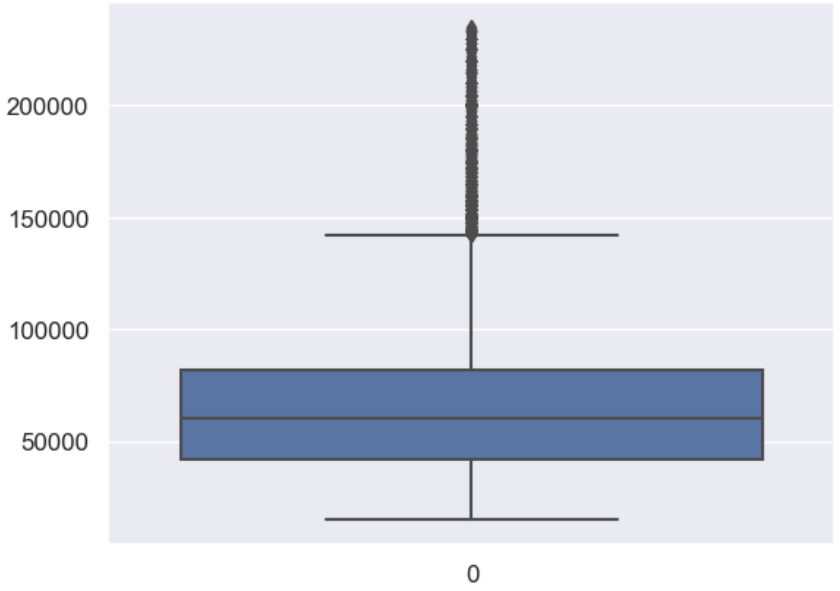
findings

#1. Average salary is 69407

#2. Maximum salary is 6000000 and minimum is 4000

#3. Most applicants have salary range between 41000 and 83000

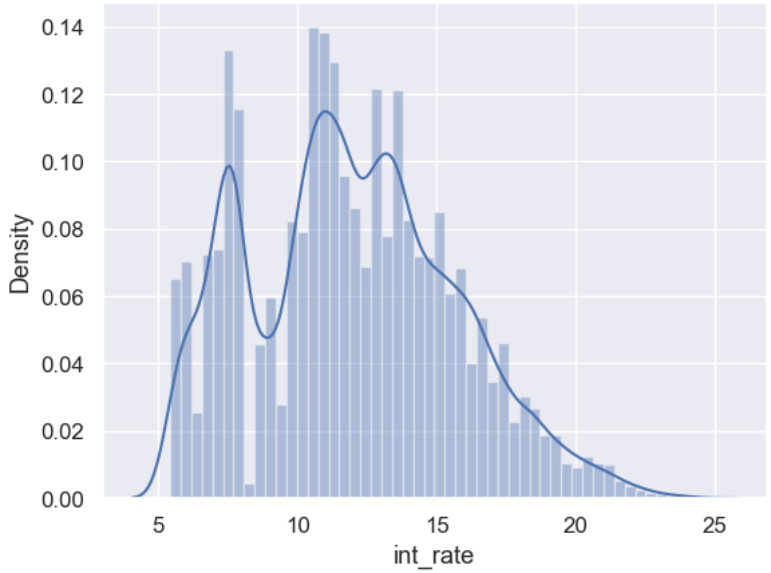


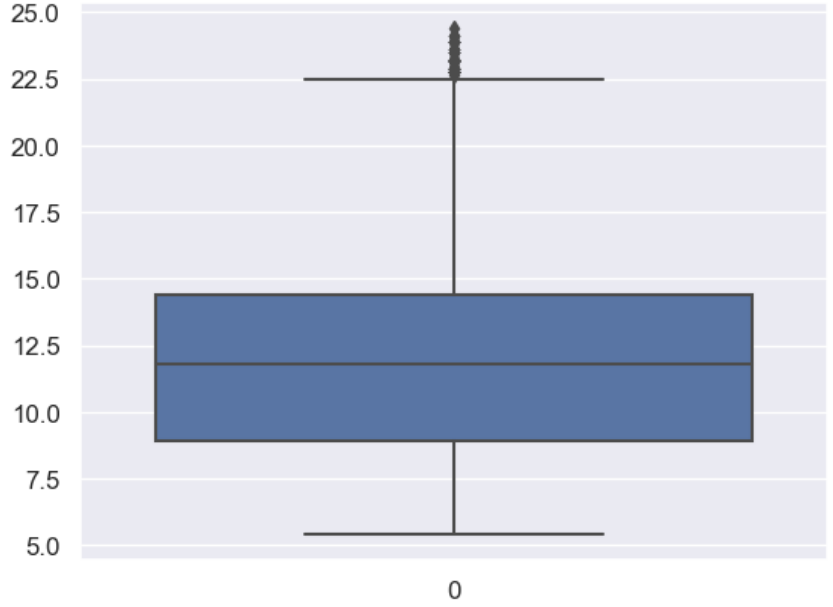


Int_Rate

findings

- Average interest rate is 12%(approximately)
- Maximum interest rate is 24.4% and minimum is 5.42%
- Most loans have interest rate between 8.94% and 14.41%
- Bimodal, i.e likely two sets of interest rates
- No evident outliers





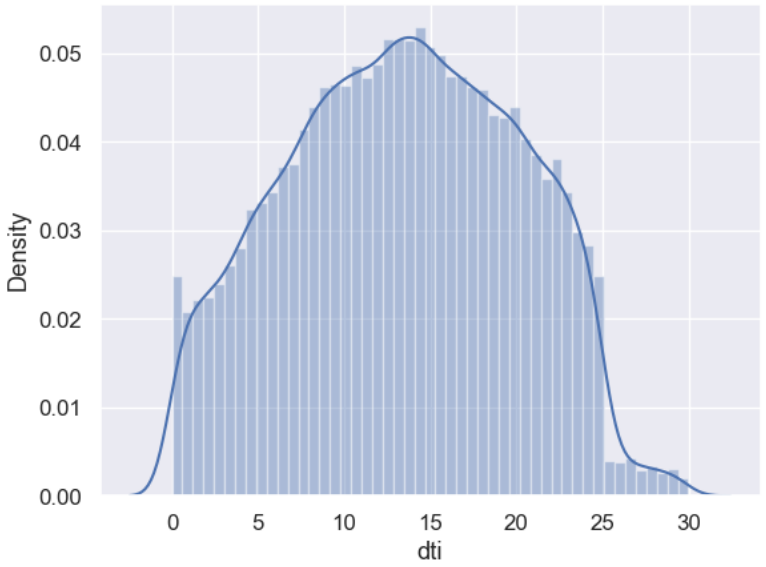
Ownership

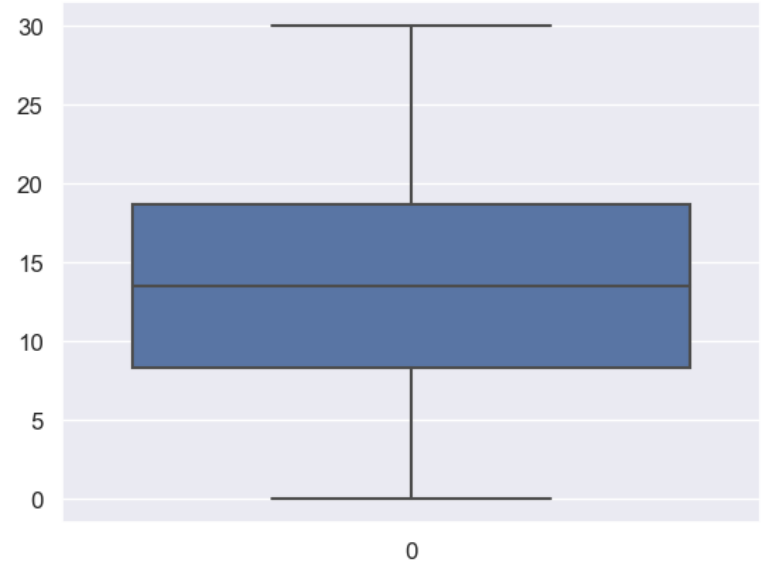
Home Ownership	Percentage
RENT	48.12
MORTGAGE	44.36
OWN	7.26
OTHER	0.26
NONE	0.01

#Findings
#1. Most people who take loans are staying at rented houses(48%) or have mortgaged their houses(44%).
#2. People with no houses generally doesn't take any loans(0.01%).

Debt to income ratio(DTI)

count	36762.000000
mean	13.375227
std	6.638618
min	0.000000
25%	8.290000
50%	13.480000
75%	18.620000
max	29.990000

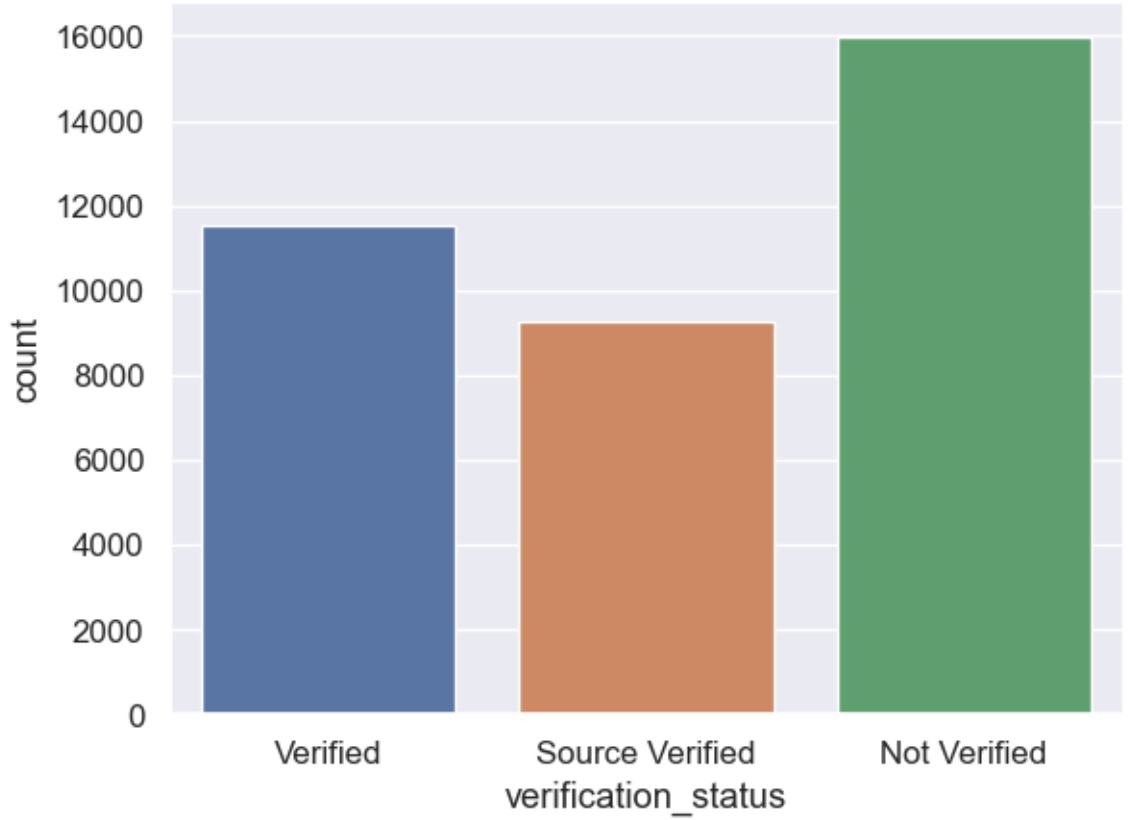




#findings
#1. Average DTI is 13(approximately)
#2. Maximum DTI is 30(approximately) and minimum is 0
#3. Most applicants have DTI between 8.29 and 18.62

Verification Status

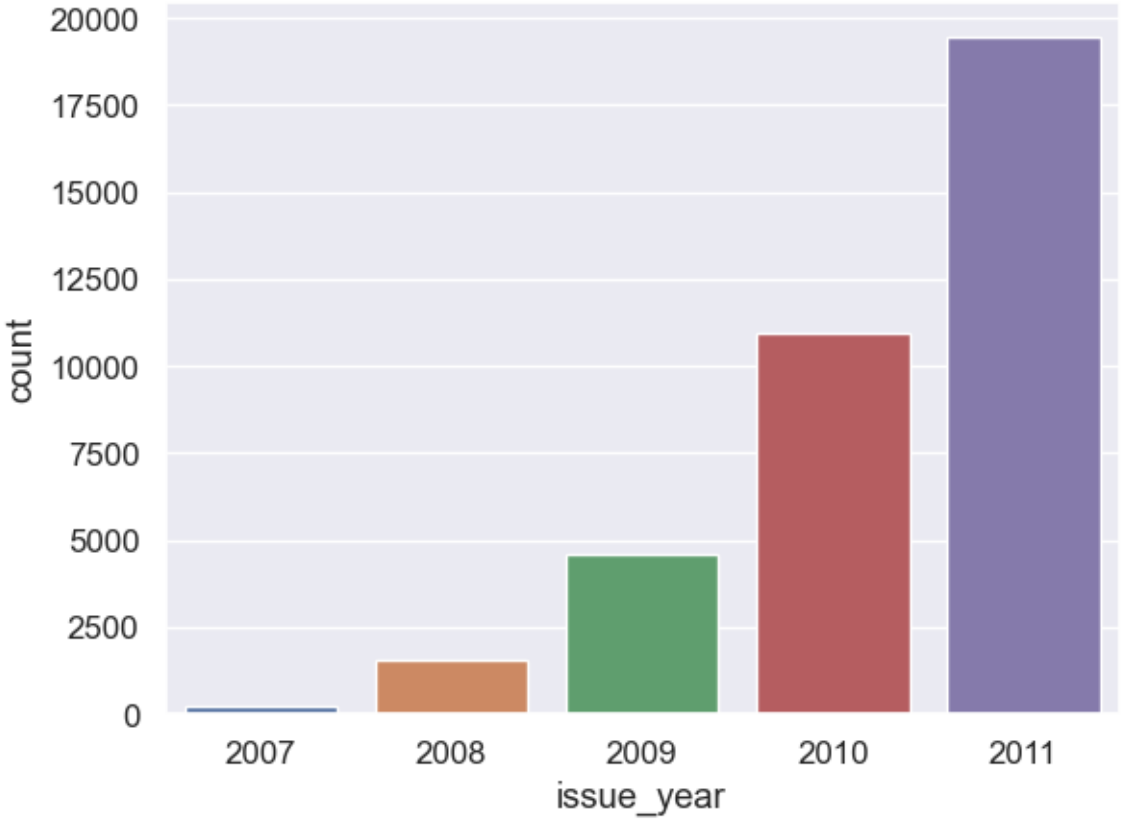
Verification Status	Percentage
Not Verified	43.46
Verified	31.37
Source Verified	25.17



#Findings
#1. Most of the loans taken are by verified people(25% + 31% = 56 %)
#2. Non verified people loans are around 43.5%

Year of issuance

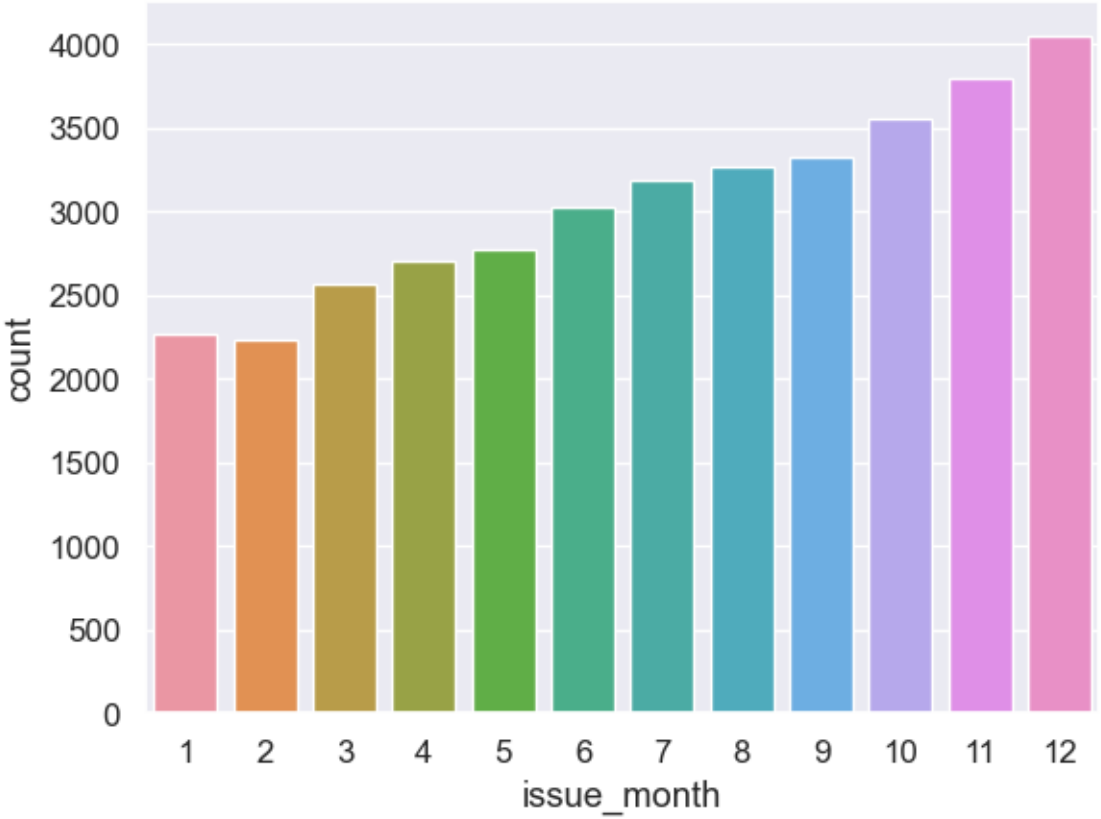
Year	Perentage
2011	52.91
2010	29.8
2009	12.47
2008	4.17
2007	0.65



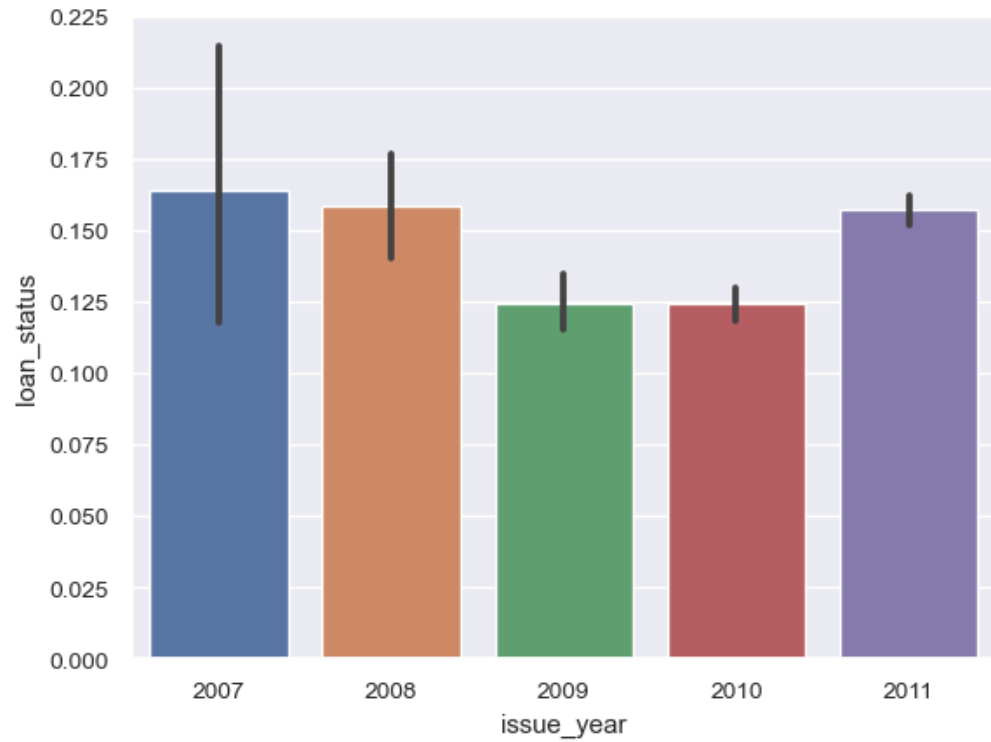
#Findings
#1. Most loans were taken in the year 2011
#2. In 2007 people took least loans(Loan interest rate might be the reason for this. Maybe in 2007-2008 loan interest rates were high)

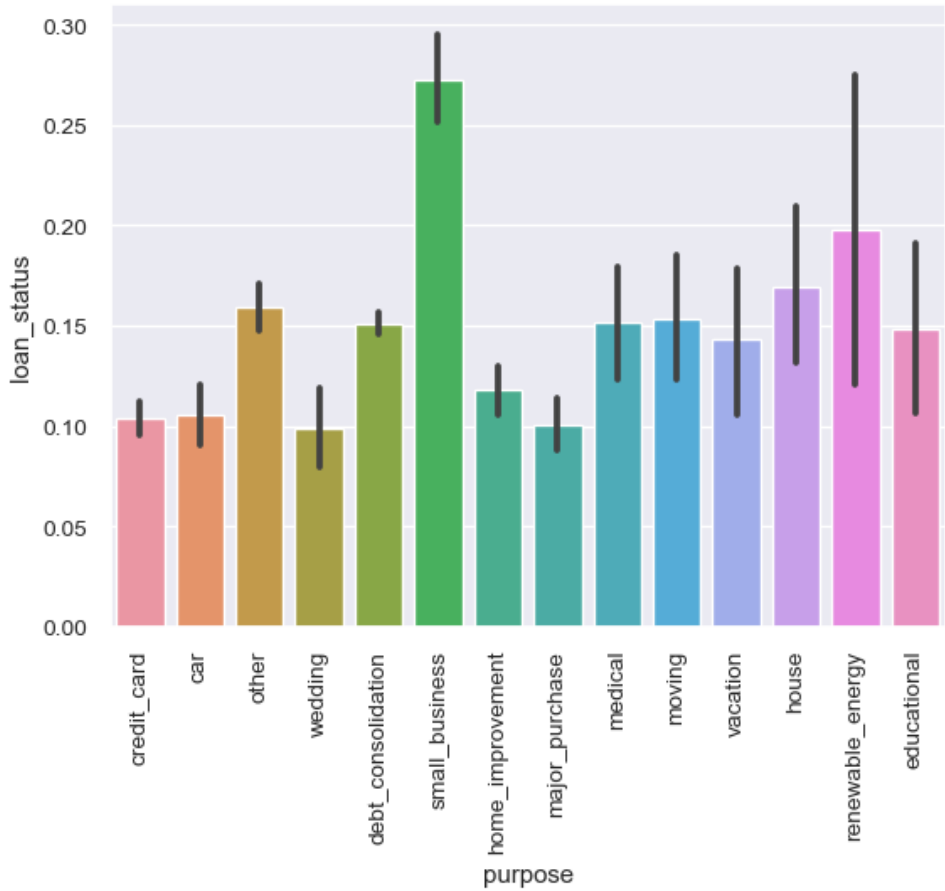
Month of issuance

Month	Percentage
12	11.02
11	10.33
10	9.66
9	9.06
8	8.88
7	8.66
6	8.24
5	7.56
4	7.36
3	6.98
1	6.18
2	6.08

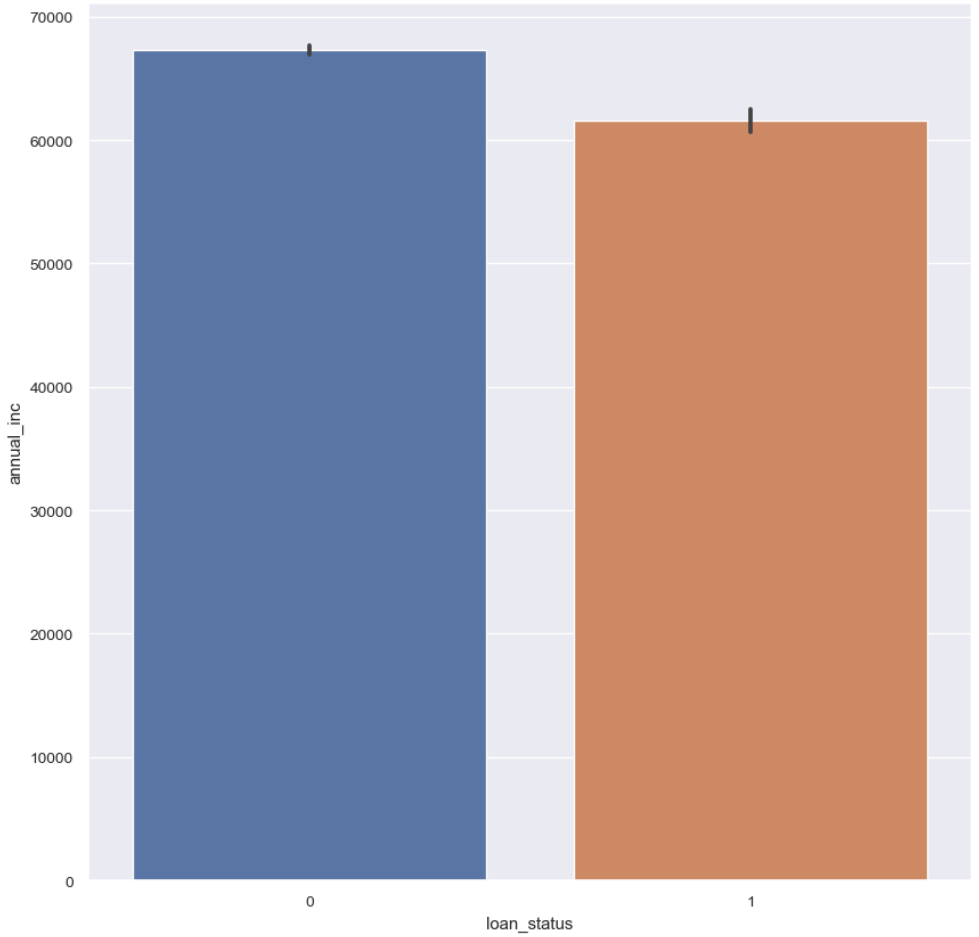


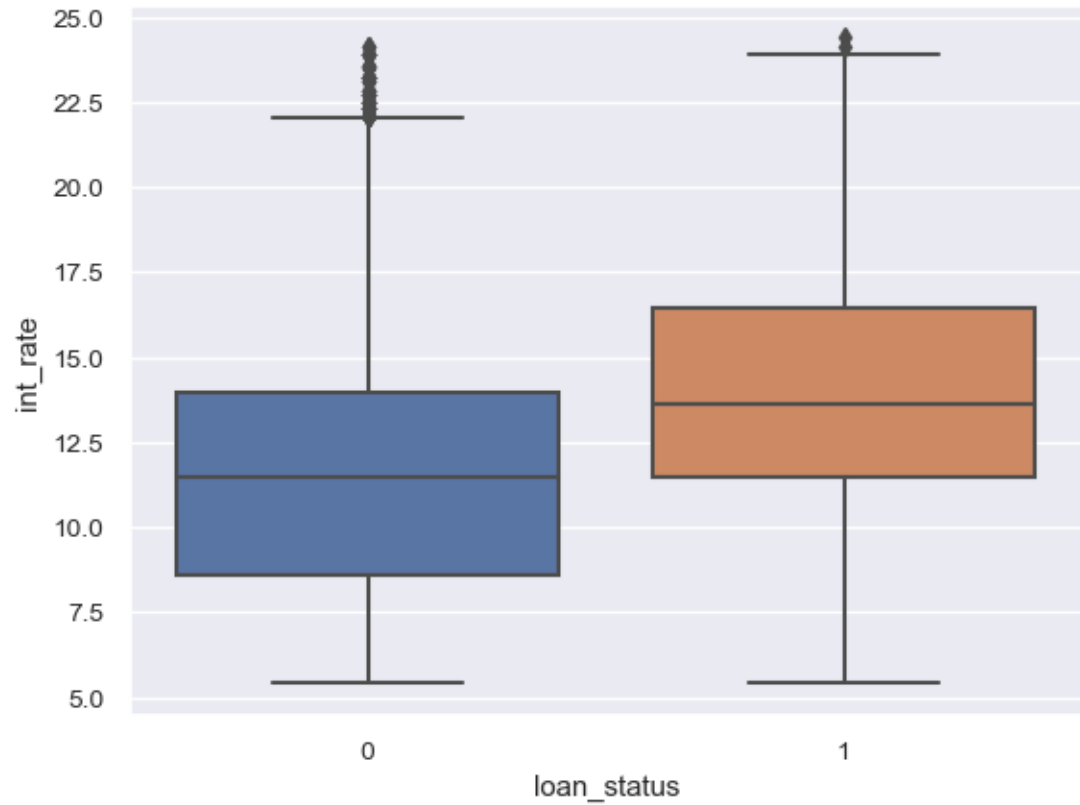
#Findings
#1. At the start of the year people generally take lesser loans
#2. At the end of the year people generally seems to take more loans.

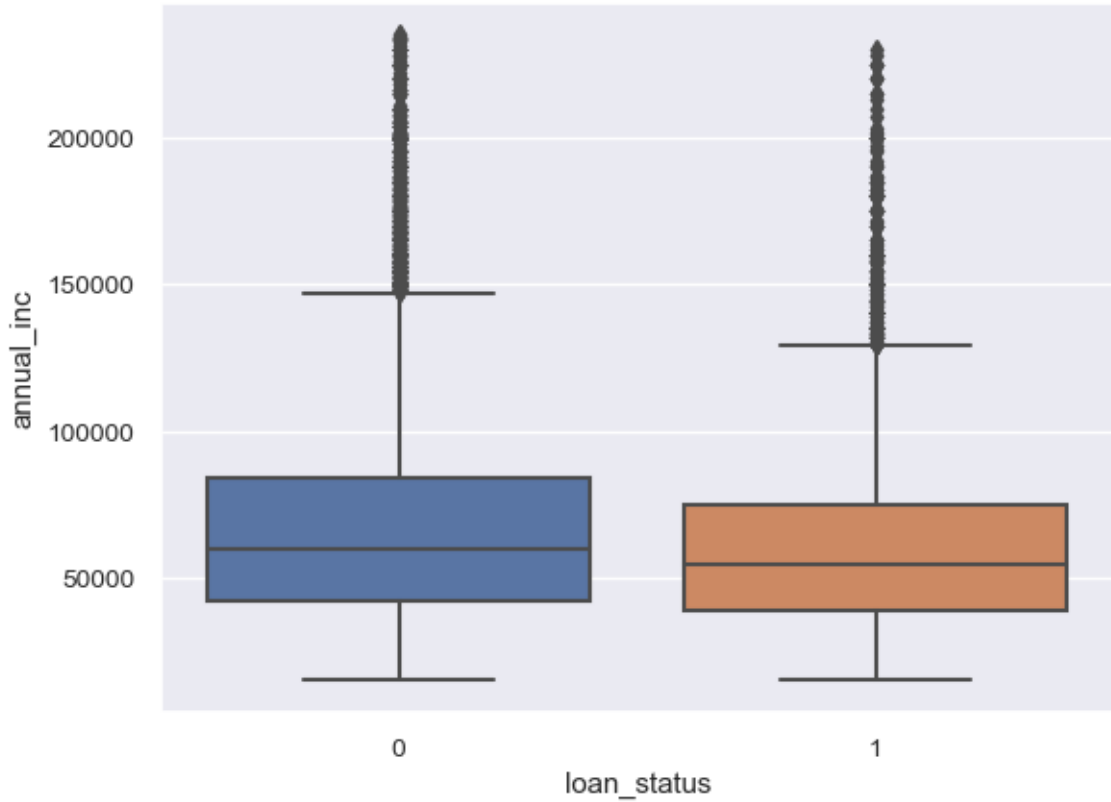
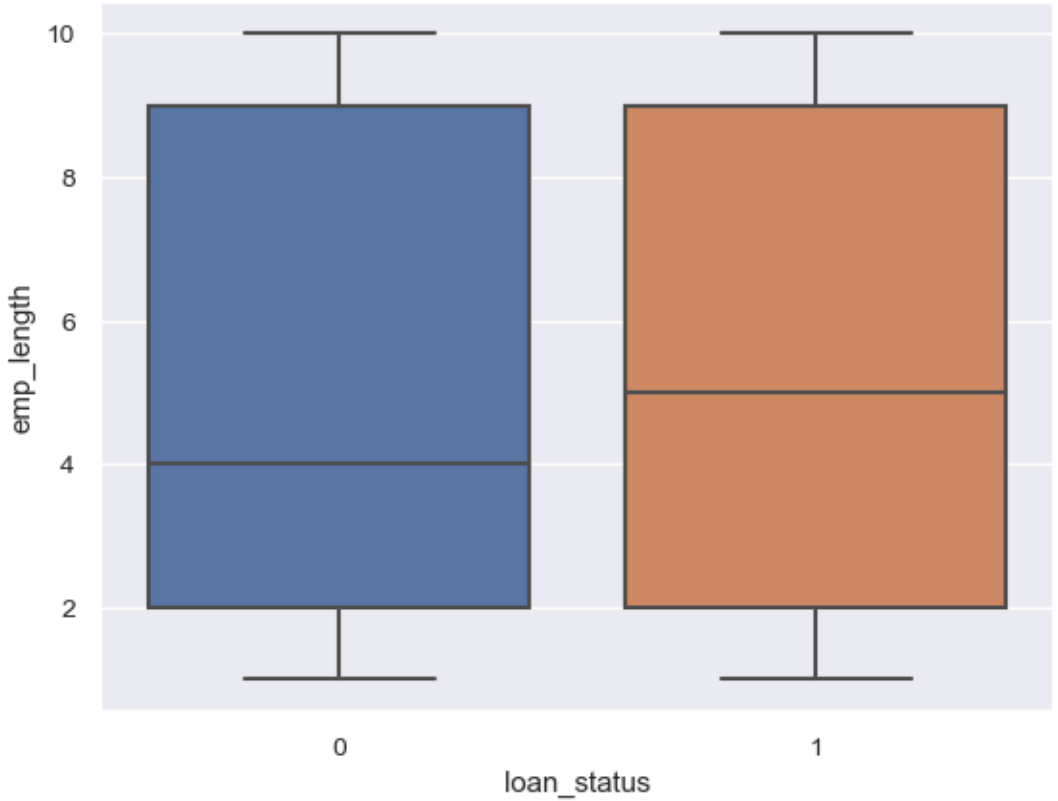

0.225
0.200
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
loan_status
2007
2008
2009
2010
2011
issue_year


0.30
0.25
0.20
0.15
0.10
0.05
0.00
loan_status
credit_card
car
other
wedding
debt_consolidation
small_business
home_improvement
major_purchase
medical
moving
vacation
house
renewable_energy
educational
purpose

Recovery rate is slightly h more when the




25.0
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0
int_rate
0
1
loan_status


200000
150000
100000
50000
annual_inc
0
1
loan_status

Employee length and loan status are not dependent