Business Objective

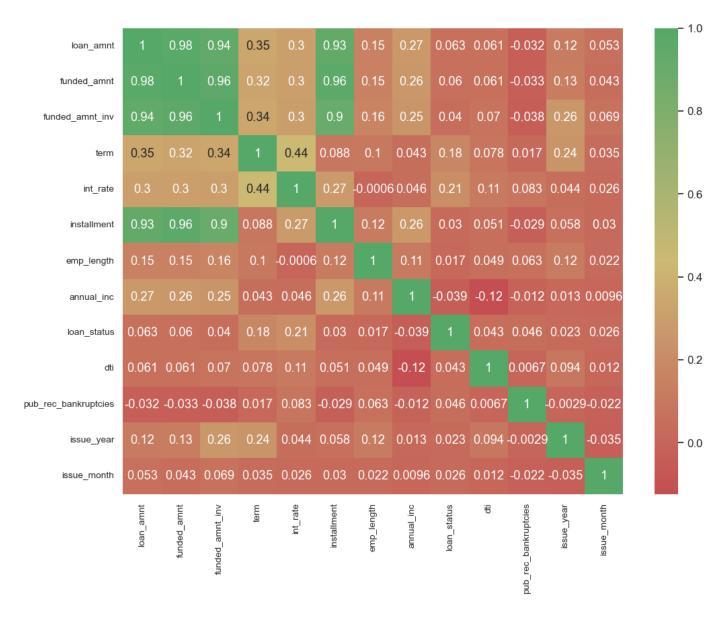
This company is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures. Borrowers can easily access lower interest rate loans through a fast online interface. Lending loans to 'risky' applicants is the largest source of financial loss (called credit loss). Credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed. In other words, borrowers who default cause the largest amount of loss to the lenders. In this case, the customers labelled as 'charged-off' are the 'defaulters'. If one is able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. Identification of such applicants using EDA is the aim of this case study.

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

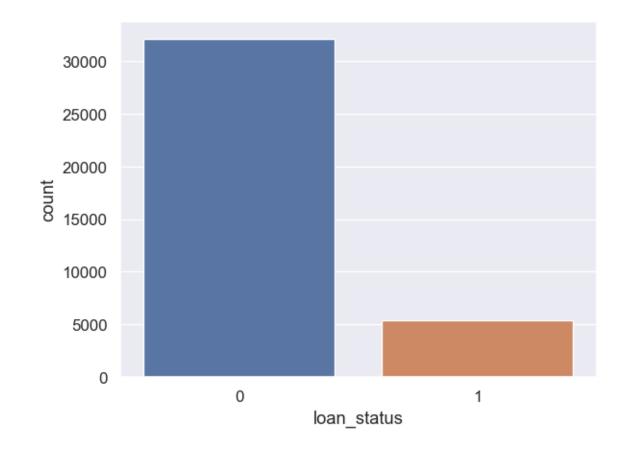
- Observations
 - Let us take sample variable as Loan status
 - The variables which are relatively higher correlation with respect to loan_status
 - Int_rate
 - term



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

11119.329986 mean

7354.098954 std

min 500.000000

25% 5500.000000

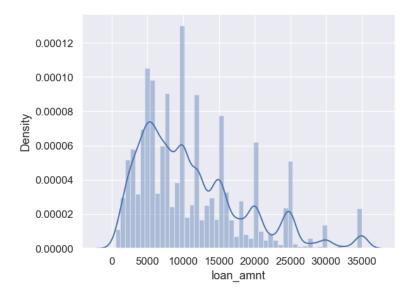
50% 10000.000000

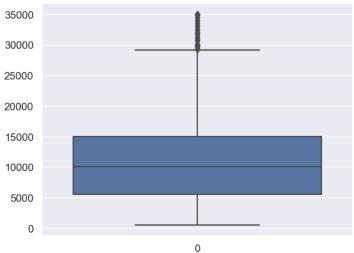
15000.000000 75%

35000.000000 max

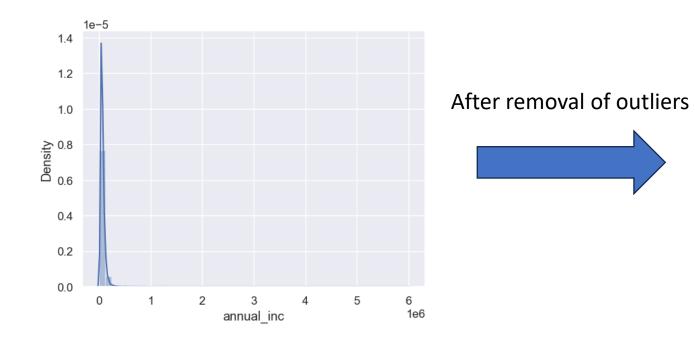
Observations

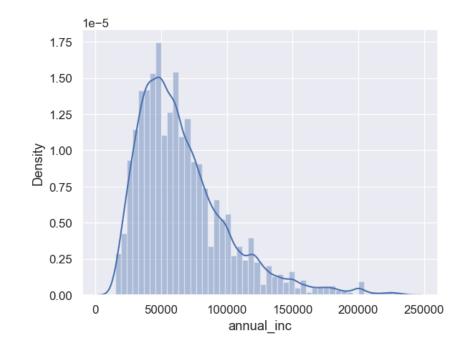
- #1. Average Loan amount is 11119
- #2. Most of the loan taken are in the range 5500 15000
- #3. Maximum loan taken is 35000 and minimum is 500





Annual Income

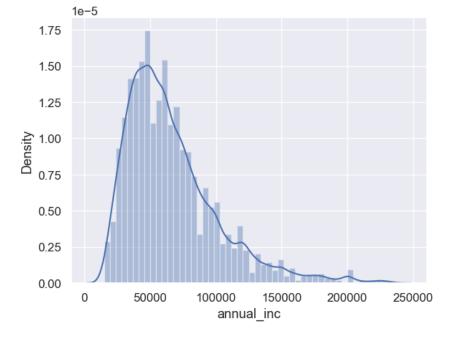


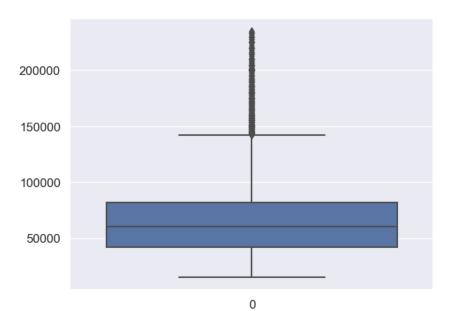


Annual Income

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

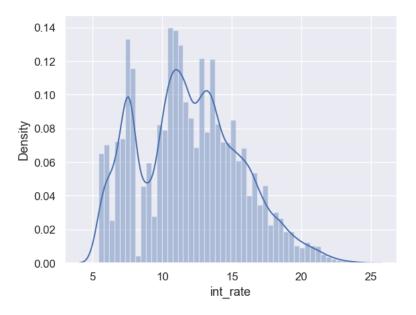


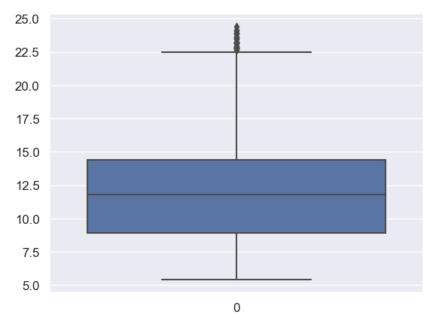


Int_Rate

findings

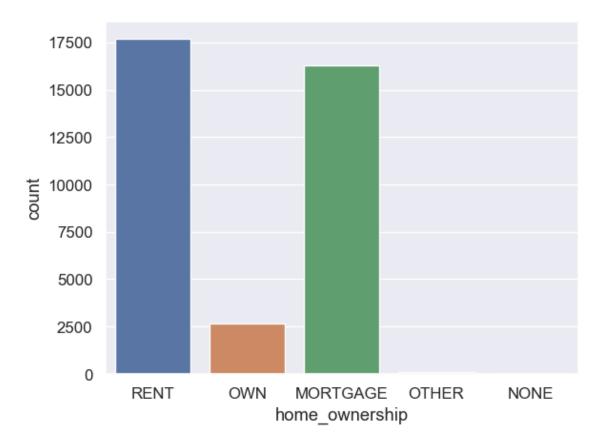
- Average interest rate is 12%(approximately)
- Maximum interest rate is 24.4% and minimum is 5.42%
- Most loans have interest rate between 8.94% and 14.41%
- Bimodal, i.e likely two sets of interest rates
- No evident outliers





Ownership

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



#Findings

loans(0.01%).

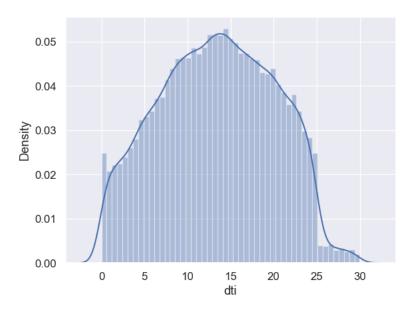
#1. Most people who take loans are staying at rented houses(48%) or have mortgaged their houses(44%). #2. People with no houses generally doesn't take any

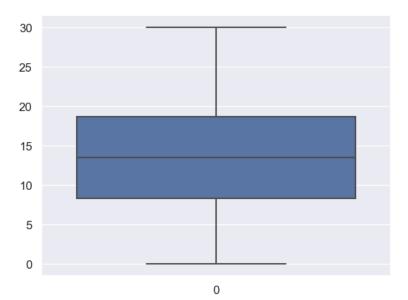
Debt to income ratio(DTI)

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

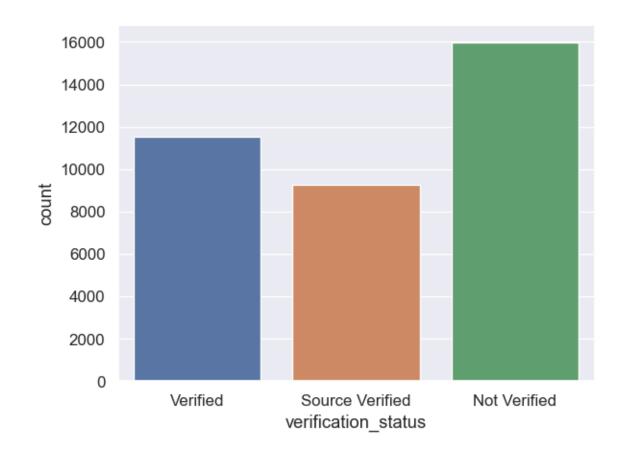
- #1. Average DTI is 13(approximately)
- #2. Maximum DTI is 30(approximately) and minimum is 0
- #3. Most applicants have DTI between 8.29 and 18.62





Verification Status

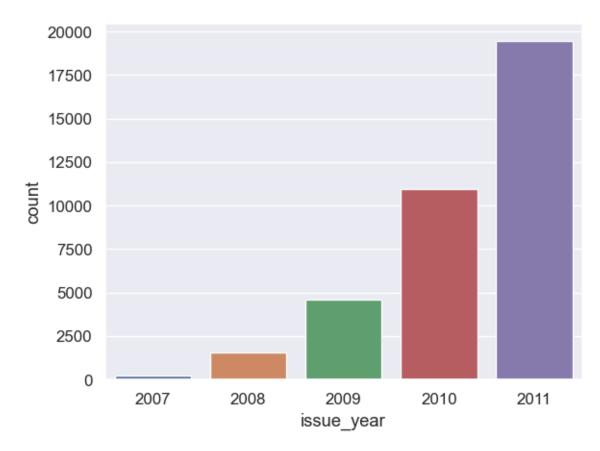
Verification Status	Percentage
Not Verified	43.46
Verified	31.37
Source Verified	25.17



- #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
20	52.91
20	29.8
20	12.47
20	008 4.17
20	0.65



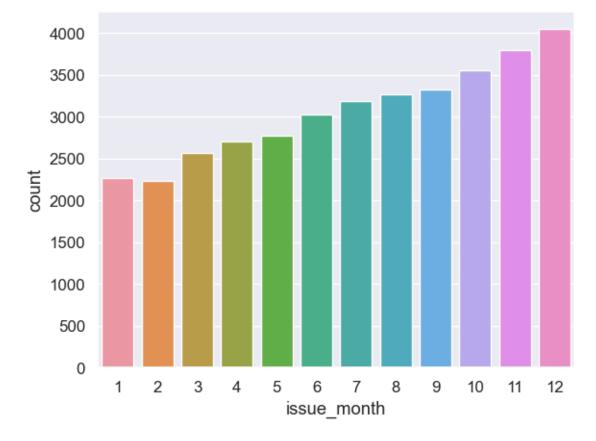
#Findings

#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

Percentage
11.02
10.33
9.66
9.06
8.88
8.66
8.24
7.56
7.36
6.98
6.18
6.08

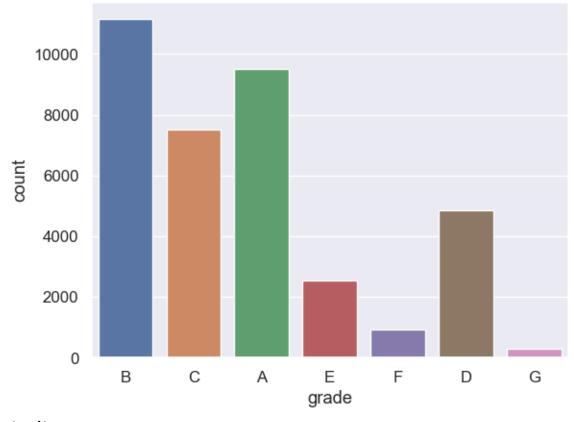


#Findings

#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.

Grade

Grade	Percentage
В	30.31
Α	25.83
С	20.4
D	13.2
Е	6.93
F	2.54
G	0.78

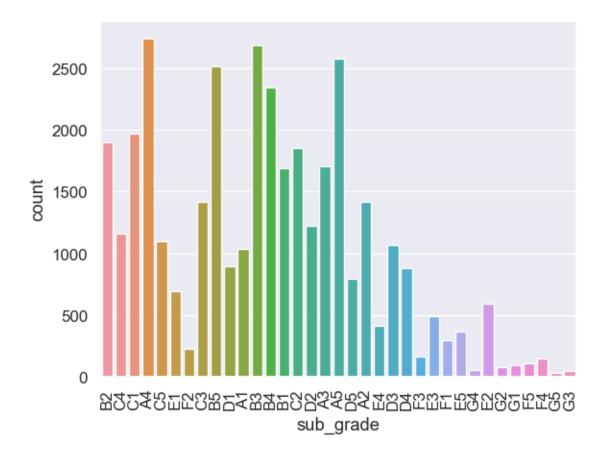


#Findings

#1. Grade B generally take more loans than any other grade followed by Grade A and Grade C respectively #2. Grade F and Grade G applicants take comparatively much lesser loans.

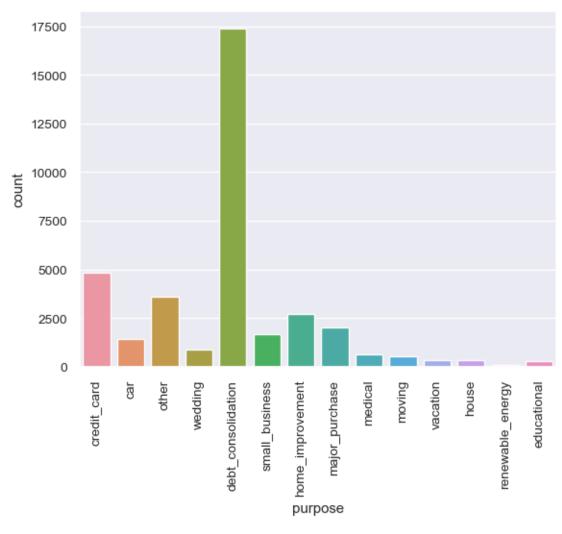
Sub Grade

- #1. Grade A4, B3, A5 take almost equal amount of loans and maximum number of loans.
- #2. Grade G5 applicants take least amount of loans.



Purpose

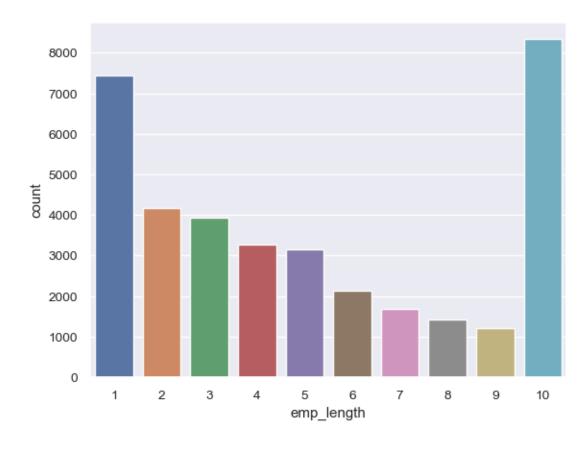
Purpose	Percentage
debt_consolidation	47.33
credit_card	13.14
other	9.81
home_improvement	7.33
major_purchase	5.52
small_business	4.52
car	3.83
wedding	2.45
medical	1.74
moving	1.44
house	0.93
vacation	0.93
educational	0.77
renewable_energy	0.25



- #1. almost 50% loans are taken to pay out existing loans.
- #2. 13% loans are taken to pay credit card balances.

Employee Experience

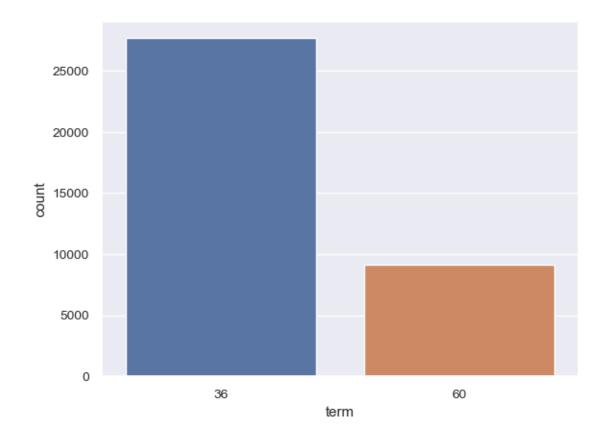
Employee Length	Percentage
10	22.69
1	20.22
2	11.34
3	10.7
4	8.92
5	8.57
6	5.82
7	4.59
8	3.85
9	3.3



- #1. People tend to take loans at the start i.e. 0-1 years of experience or after 10+ years of experience.
- #2. for 5-9 years experienced the loans taken are least.

Term

Terms in months	Percentage
36	75.22
60	24.78

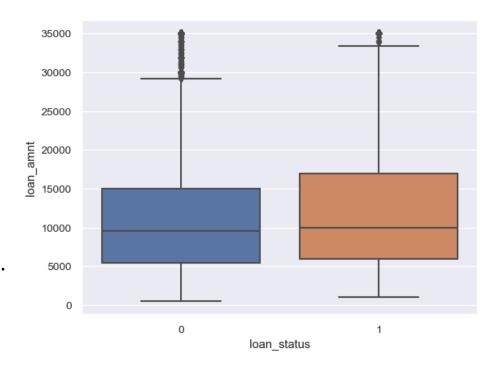


- #1. almost 75% loans are of lesser duration i.e. 36 months.
- #2. long duration loans(60 months) are around 25%

Bivariate Analysis

Loan Status vs Loan Amount

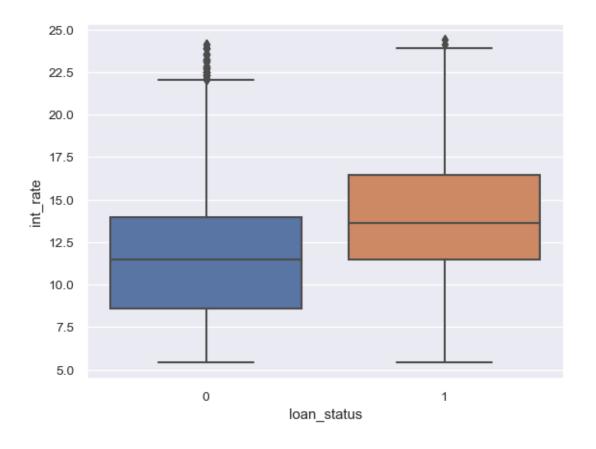
- #1. Average loan amount of defaulter is slightly high
- #2. There is no major difference in loan amounts of defaulters.



Loan Status vs Int Rate

#Findings

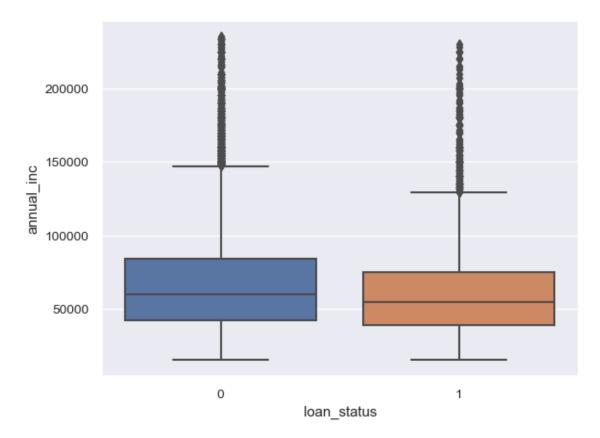
#1. Applicants with high interest rate seems to default more than lower interest rates.



Loan Status vs Annual Income

#Findings

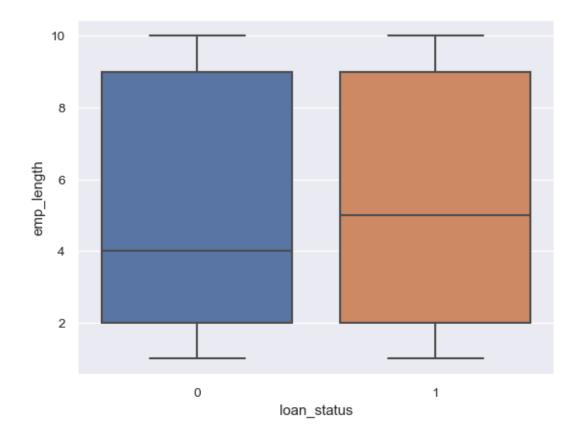
#1. Applicants with higher salary seem to default a little lesser.



Loan Status vs Emp Length

#Findings

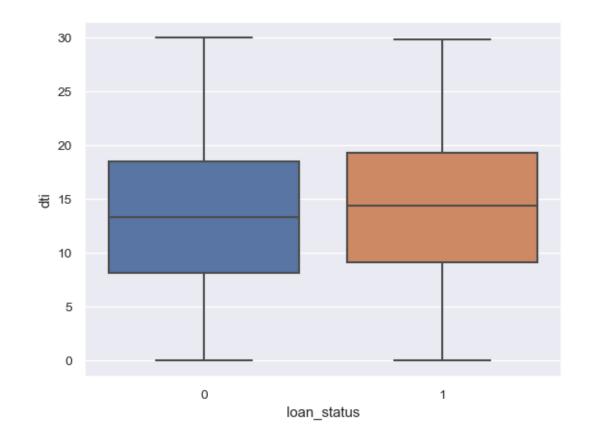
#1. Applicants with longer employment length seem to default higher.



Loan Status vs Emp Length

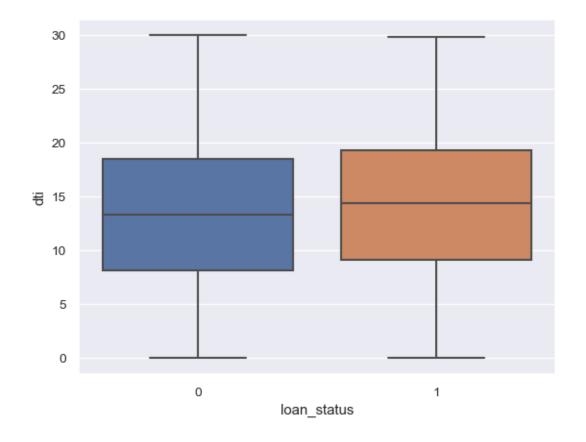
#Findings

#1. Applicants with longer employment length seem to default higher.



Loan Status vs dti

#Findings #1. applicants with higher dti seems to default a little more.

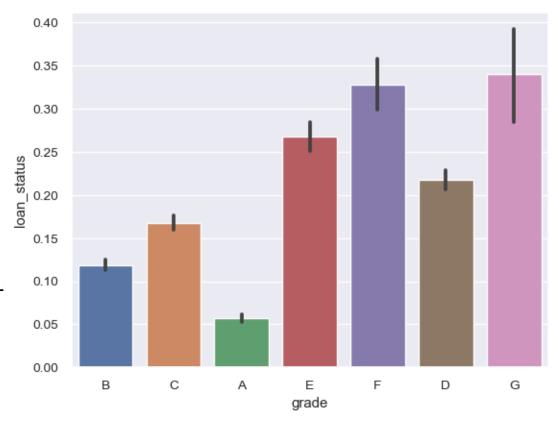


Loan Status vs Grade

#Findings

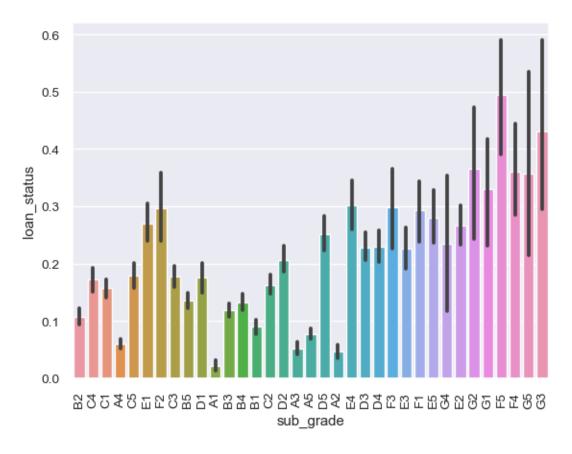
#Clearly, as the grade of loan goes from A to G, the increases.

#This is expected because the grade is decided by L based on the riskiness of the loan.



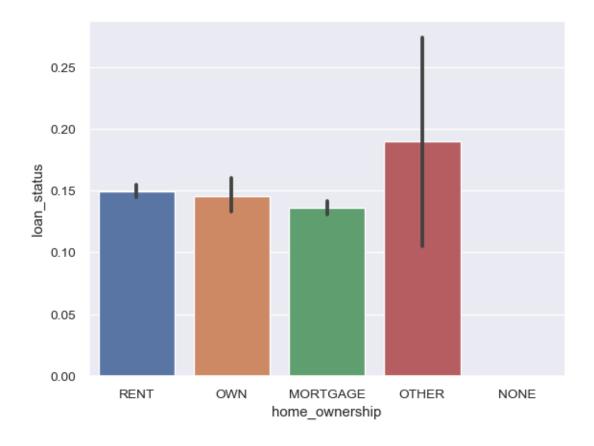
Loan Status vs Grade

#Findings #as expected - A1 is better than A2 better than A3 and so on



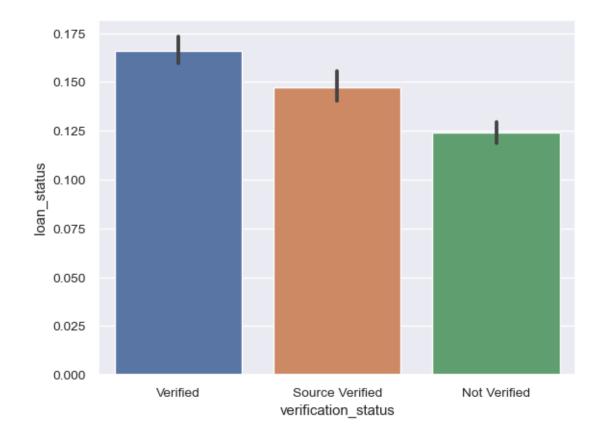
Loan Status Vs Home Ownership

#Findings
#home ownership: not a great discriminator



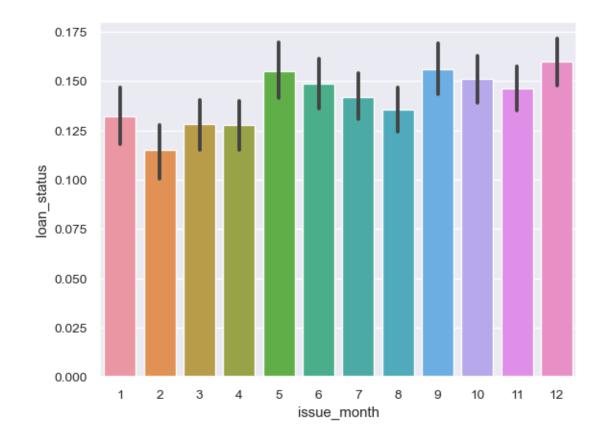
Loan Status Vs Verification Status

#Findings
#surprisingly, verified loans default
more than not verified



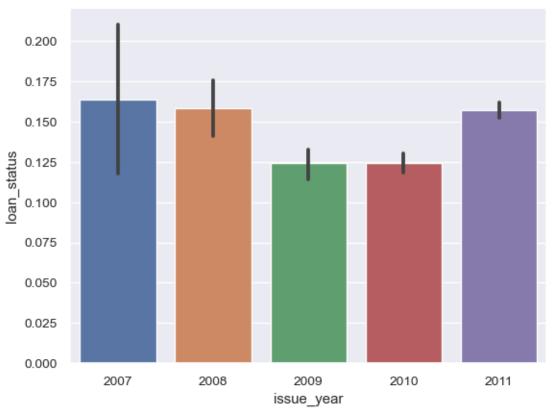
Loan Status Vs Issue Month

#Findings #1. not much variation across months



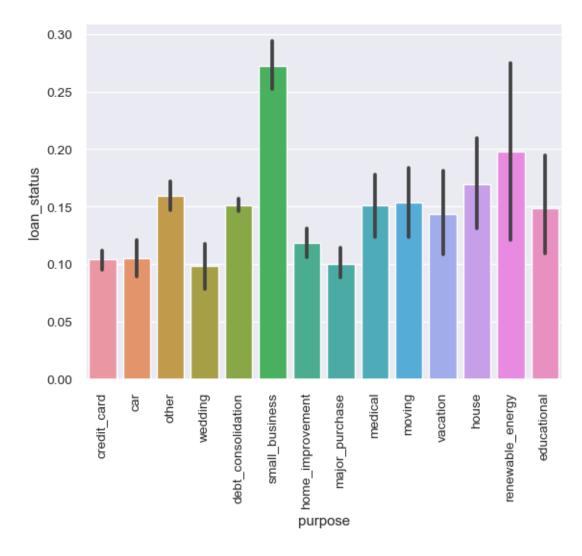
Loan Status Vs Issue Year

#Findings
#the default rate had suddenly increased in 2011, in spite of reducing from 2008 till 2010



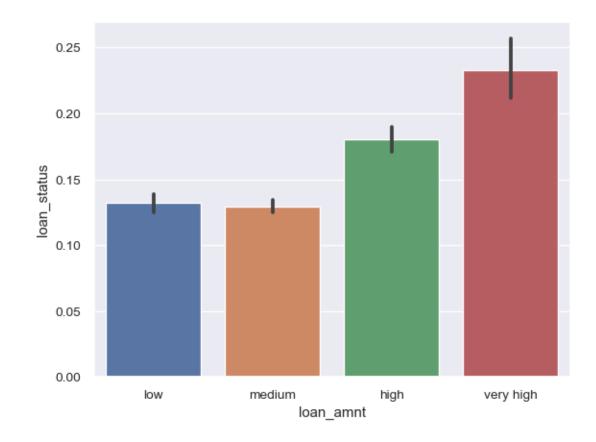
Loan Status Vs Purpose

#Findings
#Small business tends to default more



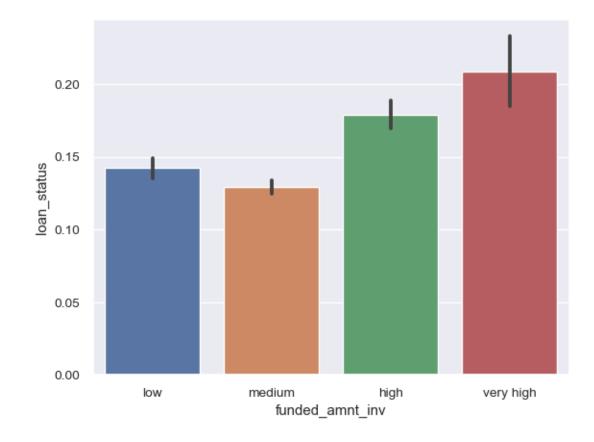
Loan Status Vs Ioan amount

#Findings
#higher the loan amount, higher the default rate



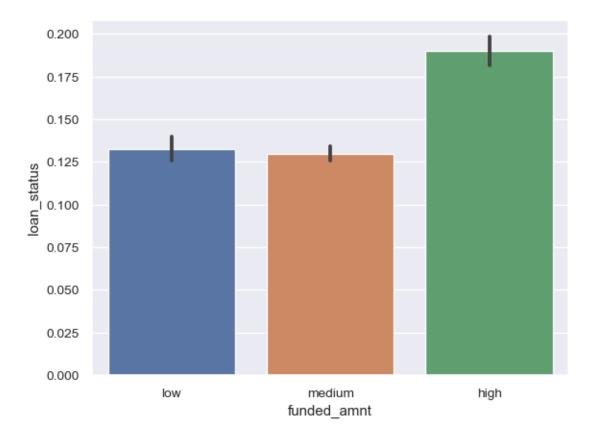
Loan Status Vs Funded_amnt_inv

#Findings
#higher the funded amount by investor,
higher the default rate
#a little dip can be seen from low to medium though



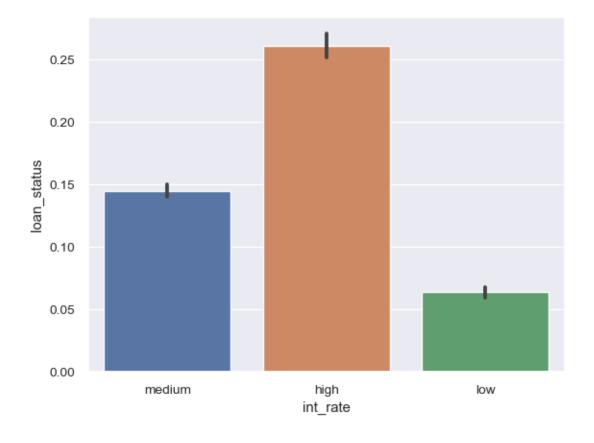
Loan Status Vs Funded_amnt

#Findings
#higher the funded amount, higher the default rate
#a little dip can be seen from low to medium though



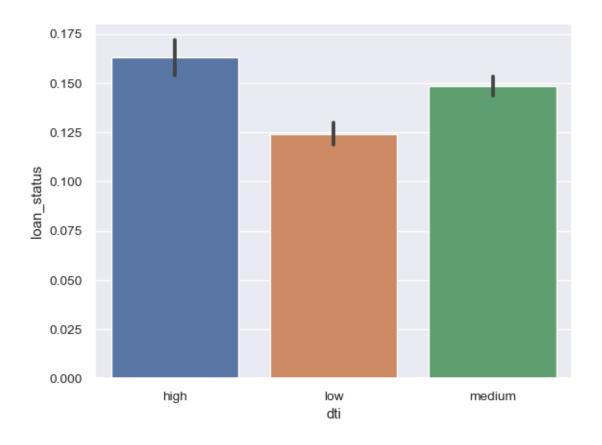
Loan Status Vs Int_Rate

#Findings
#high interest rates default more, as expected



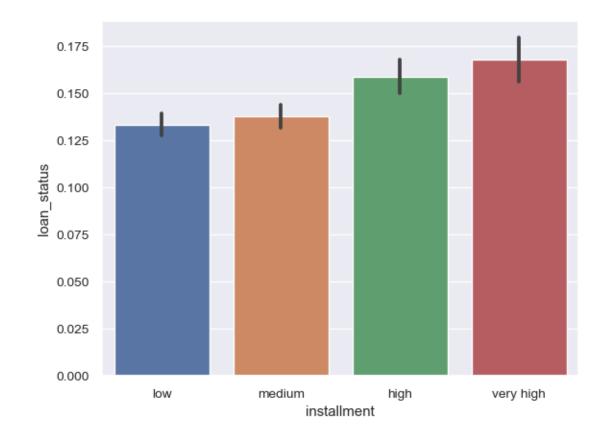
Loan Status Vs DTI

#Findings
#high dti translates into higher default rates, as
expected



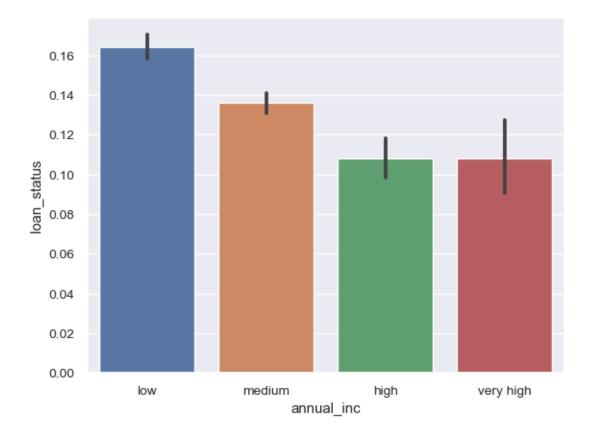
Loan Status Vs Installments

#Findings
#the higher the installment amount,
the higher the default rate



Loan Status Vs Annual Income

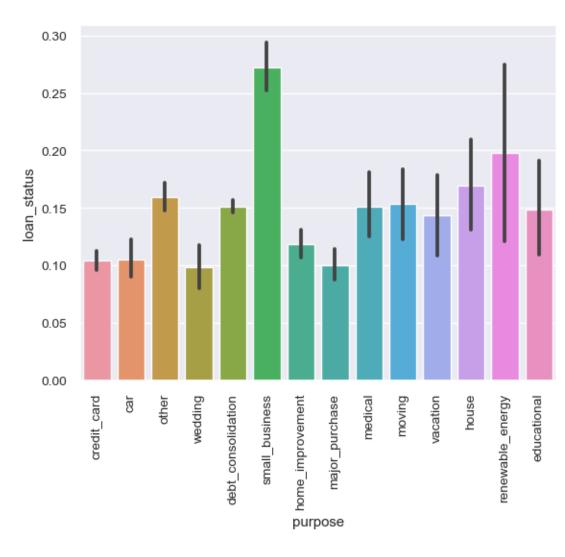
#Findings #lower the annual income, higher the default rate

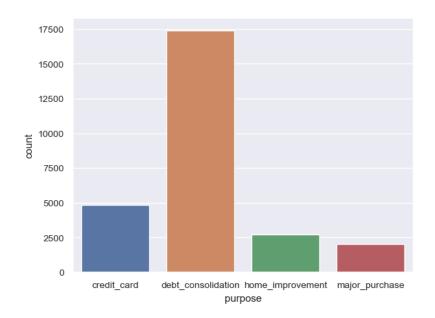


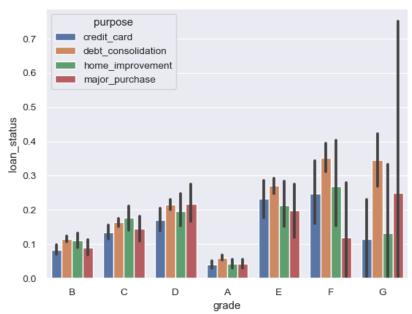
Segment Analysis

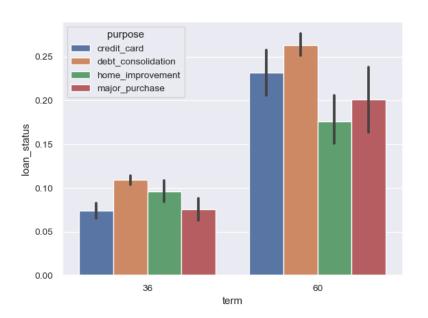
Segmented univariate analysis

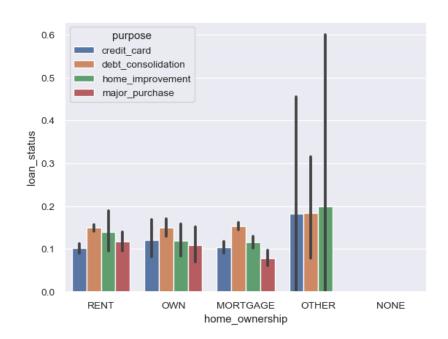
#We have now compared the default rates across various variables, and some of the important predictors are purpose of the loan, interest rate, annual income, grade etc.
#In the credit industry, one of the most important factors affecting default is the purpose of the loan - home loans perform differently than credit cards, credit cards are very different from debt consolidation loans etc.
#This comes from business understanding, though let's again have a look at the default rates across the purpose of the loan.

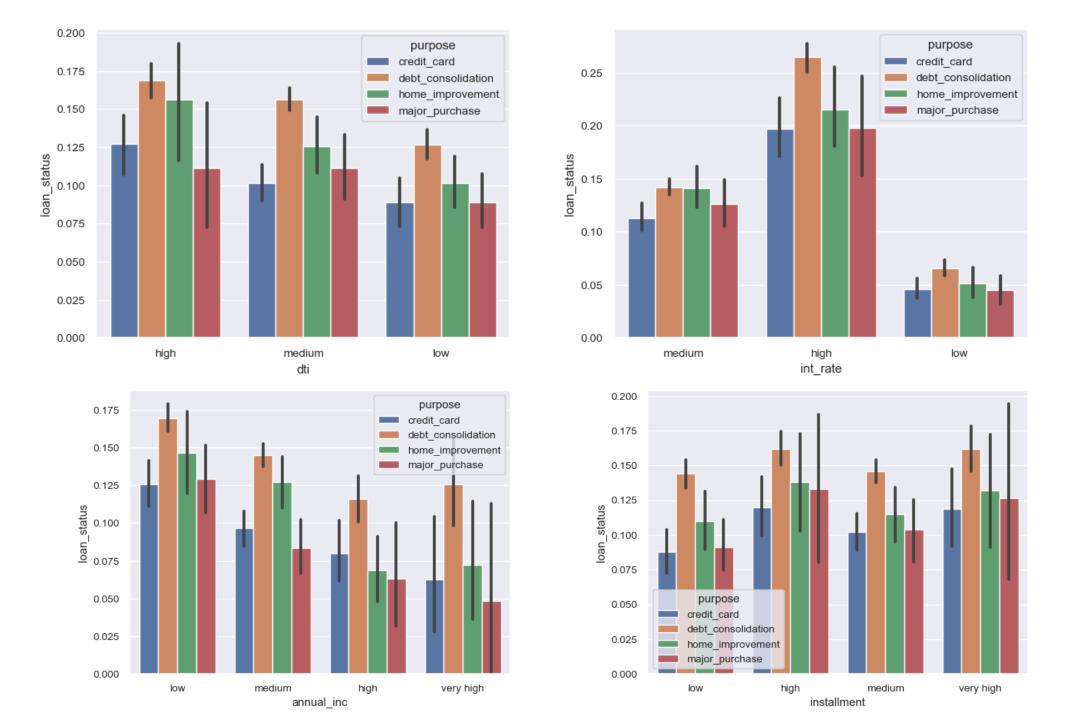












Conclusion

- Higher the interest rate higher charged off ratio
- Higher the annual income higher the loan amount slightly
- Interest rate is increasing with loan amount increase this results in high charged off.
- Mostly applicants take loan for debt consolidation and hence the highest default rate is in the same category
- Surprisingly verified loans has more default rate than unverified default rates.