Problem Statement

Using the provided loan Data set we need to give recommendations about the factors which can help in identifying if a person who has applied for loan will defaulted or fully paid. So our recommendations will help company in Approving or Rejecting loans.

Data Understanding

This loan data set contains data of people who took loan in Past or having a loan currently and it shows the status of their loan i.e., Fully Paid, Current , Charged Off

Loans payment is divided into 3 categories:

1. Fully Paid loans are good loans as principal and interests were fully recovered.
2. Charged Off loans are bad loans and amount given on loan could not be recovered as member did not pay the loan amount.
3. Current Loans are on going loans which do not given any insight about loan will be recovered or not.. so for us its neither good/bad loan.

There are other attributes present in dataset.

loan_amount : amount applied for loan

funded_amount : approved amount by lending club.

funded_amnt_inv : amount given by Investor.

emp_length : Employment Length of borrower

annual_inc : Annual Income of borrower.

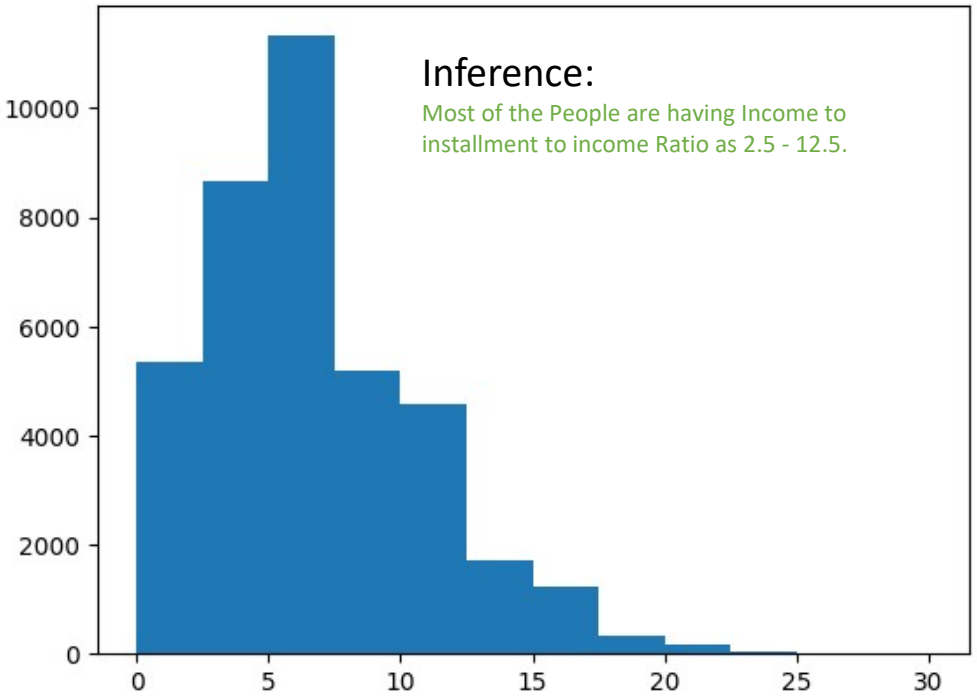
Purpose : Purpose of loan taken

Data Cleaning and Manipulation

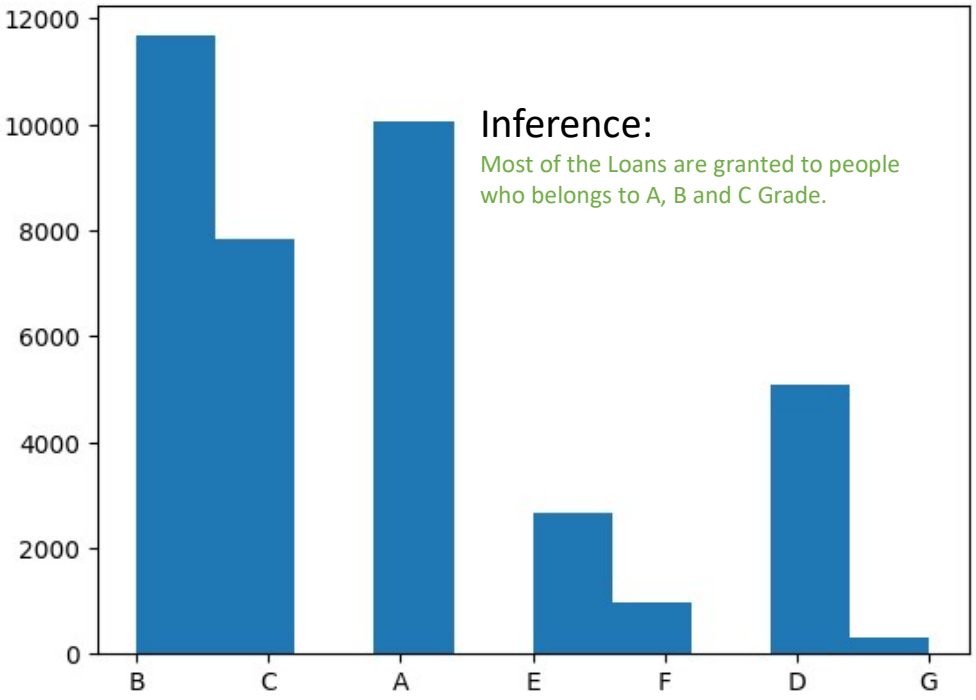
1. There are many behaviour variables which will not be present when a person applies for loans. For Example: num_tl_90g_dpd_24m.
2. Data type of some atributes were changed to int for better analysis
3. Outliers were removed from annual_inc to have better view on annual income
4. New column was added as Monthly_Installment_IncomeRatio for computing installment vs monthly income ration

Data Analysis

Uni-Variate Analysis of Monthly_Installment_IncomeRatio.
Monthly Instalment Ratio is defined as ratio of installment and monthly income

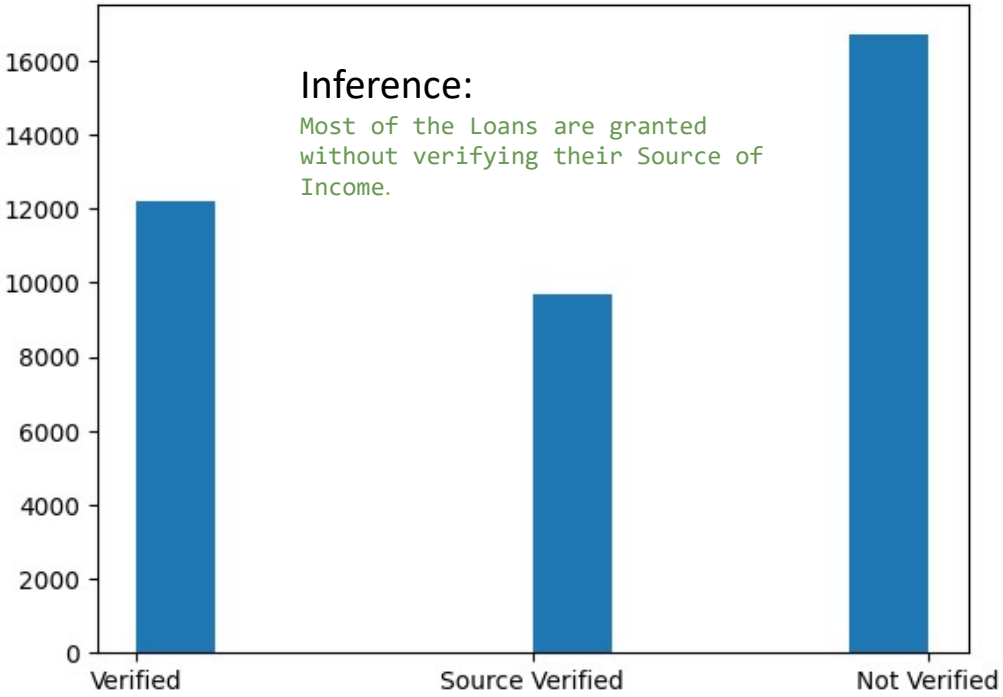


Uni-Variate Analysis of Grade.
Grade is defined by lending club

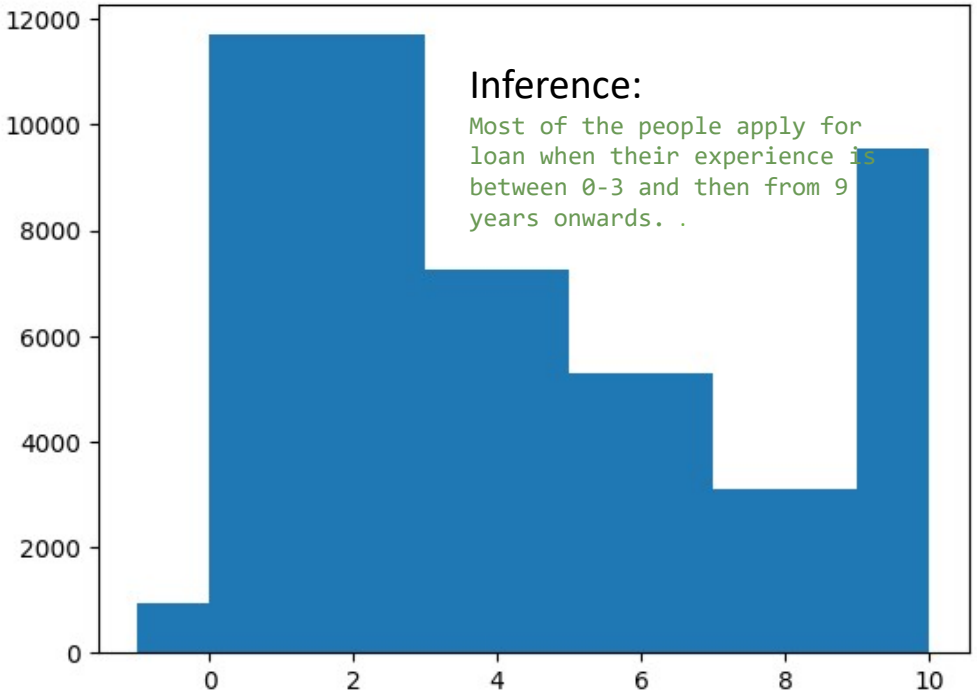


Data Analysis

Uni-Variate Analysis of verification_status
verification_status says if the source of income is verified.



Uni-Variate Analysis of Monthly_Installment_IncomeRatio.
Monthly Instalment Ratio is defined as ratio of installment and monthly income



Data Analysis

Uni-Variate Analysis of purpose
Purpose is reason for taking loan

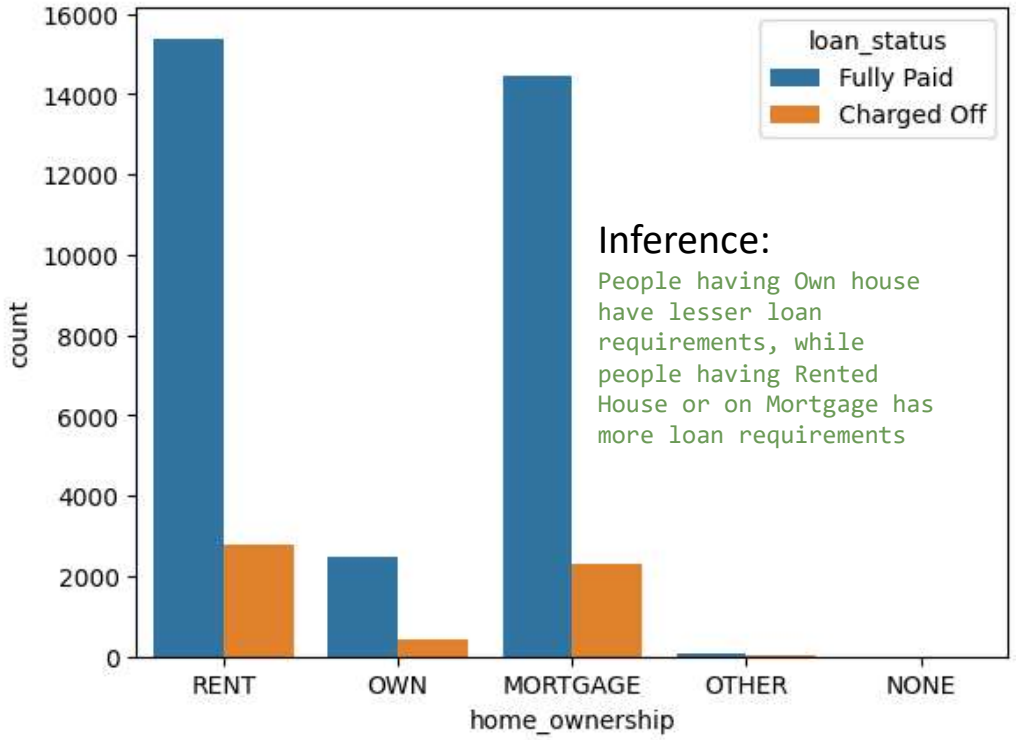
debt_consolidation	17811
credit_card	4967
other	3767
home_improvement	2791
major_purchase	2105
small_business	1707
car	1465
wedding	915
medical	667
moving	557
vacation	366
house	357
educational	293
renewable_energy	98

Inference:

Most if the People took loan to clear the debts. and then Credit card and for other purpose.

Data Analysis

Bi-Variate Analysis of home_ownership/loan_status



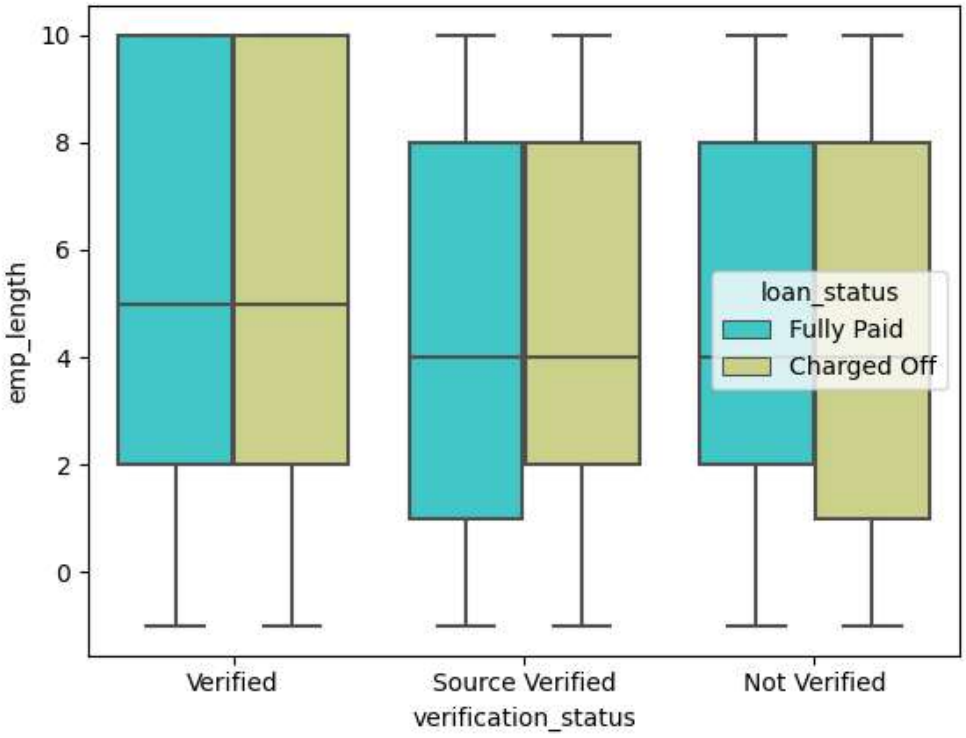
Bi-Variate Analysis of term /loan_status

Data Analysis

Multi-Variate Analysis of
emp_length/verification_status/loan_status

Inference:

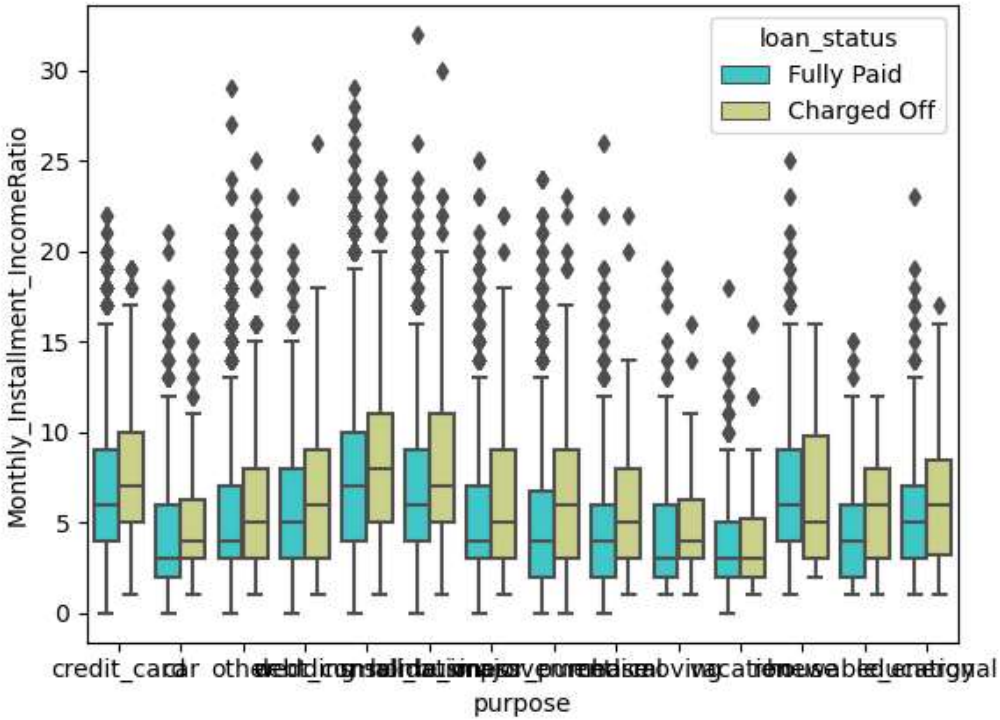
Incase of Verified Loans from Either source, whether its LC or Source, chances of defulting does not change
But it changes when source of income is not verified and Employment length is lesser



Multi-Variate Analysis of Monthly_Installment_IncomeRatio
/purpose/loan_status

Inference:

Depending on purpose of loan, the chances of loan being Charged Off or Bad loan increases

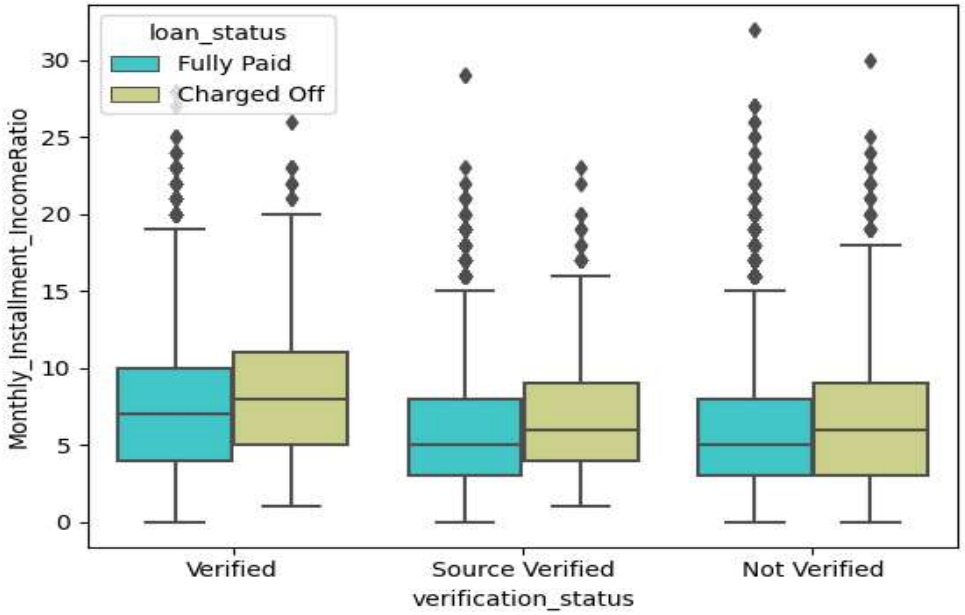


Data Analysis

Multi-Variate Analysis of Monthly_Installment_IncomeRatio/verification_status/loan_status

Inference:

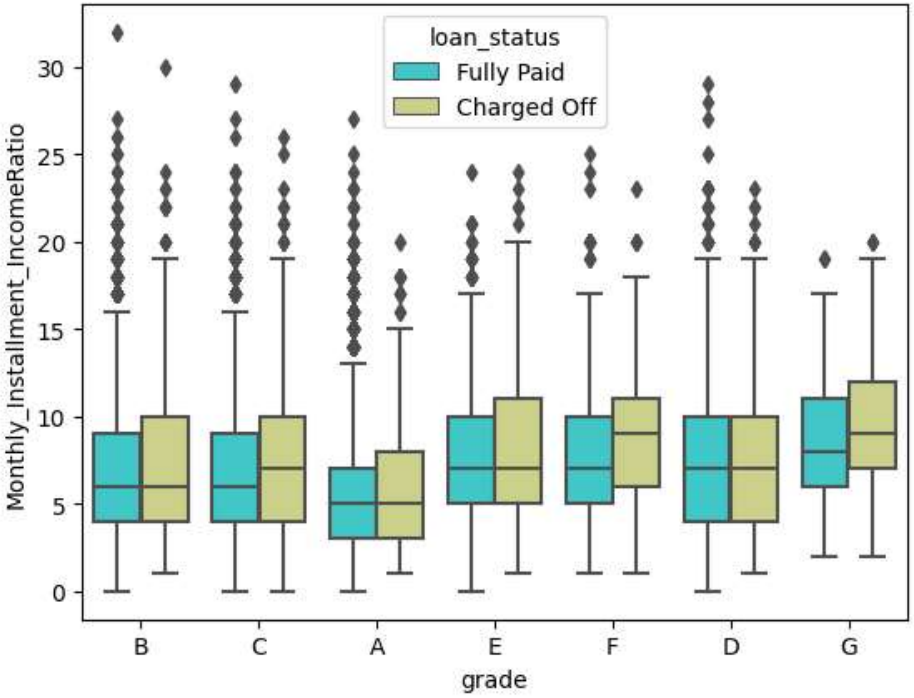
Depending on Source of Income Verification Status, the chances of loan being Charged Off or getting converted to Bad loan increases



Multi-Variate Analysis of Monthly_Installment_IncomeRatio /grade/loan_status

Inference:

Depending on Grade the loan getting converted to bad loan increases on higher side of Monthly_Installment_IncomeRatio %

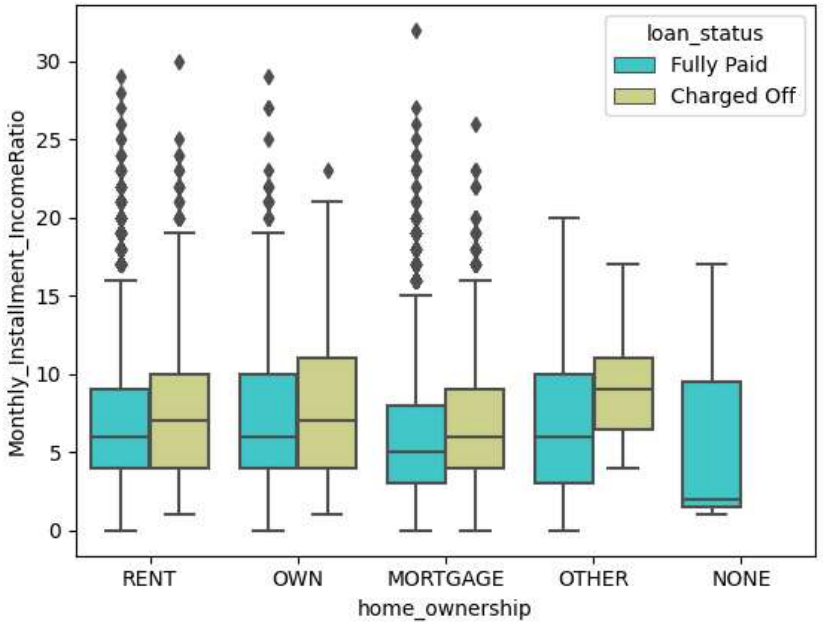


Data Analysis

Multi-Variate Analysis of Monthly_Installment_IncomeRatio/home_ownership/loan_status

Inference:

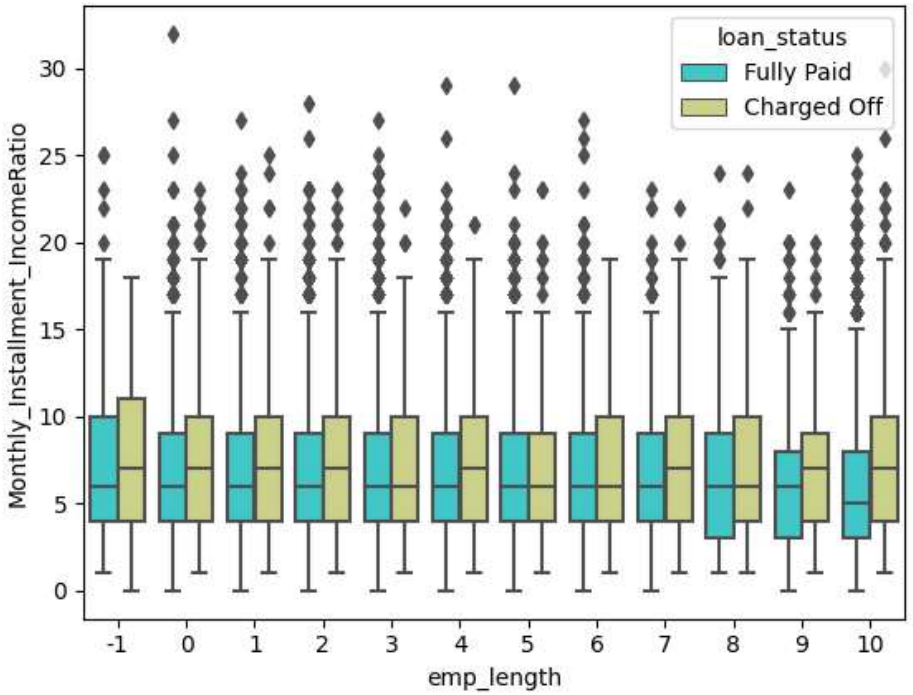
Depending on home_ownership the loan getting converted to bad loan increas on higher side of Monthly_Installment_IncomeRatio %



Multi-Variate Analysis of Monthly_Installment_IncomeRatio /emp_length/loan_status

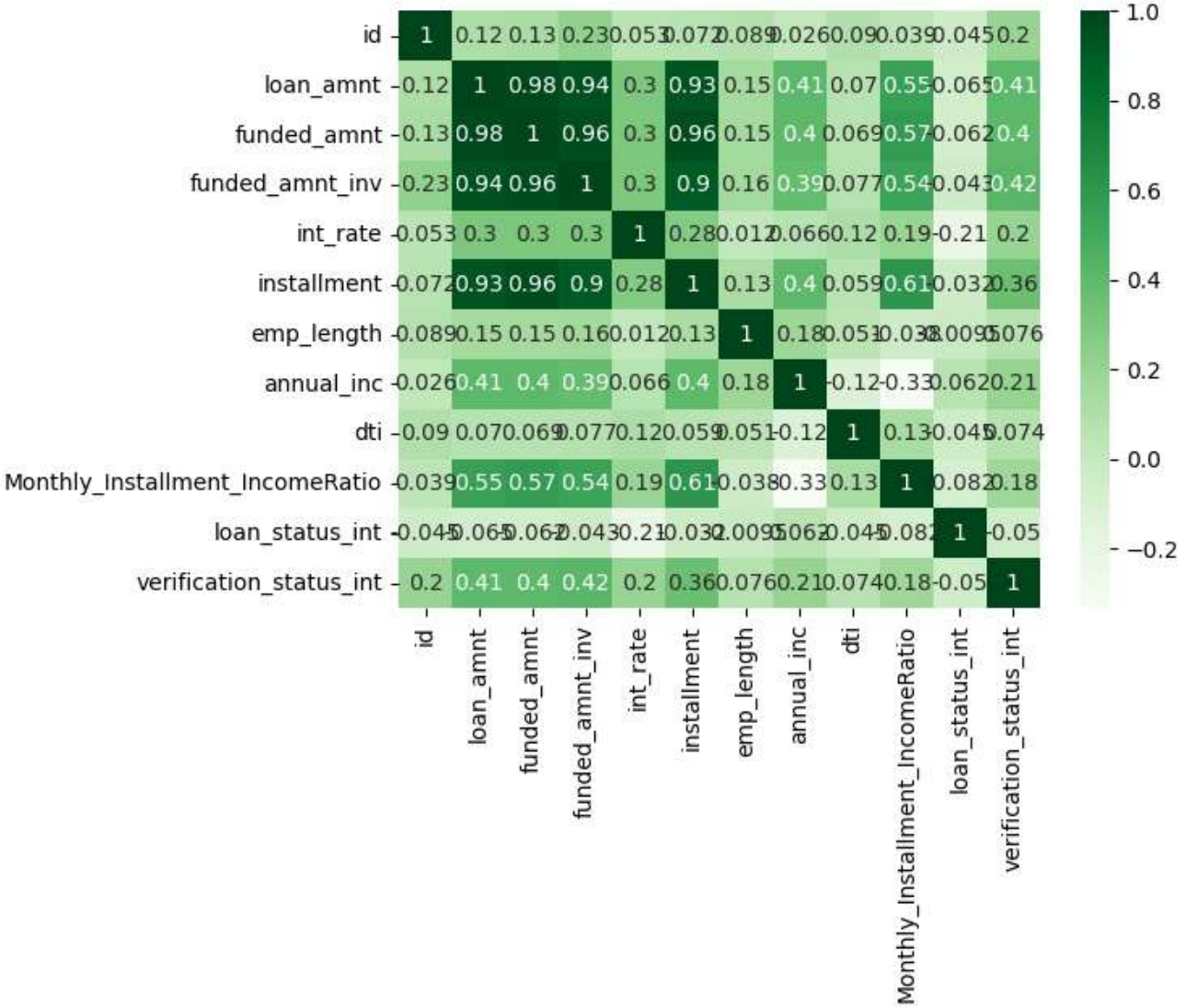
Inference:

Depending on emp_length the loan getting converted to bad loan increas on higher side of Monthly_Installment_IncomeRatio %



Data Analysis

Heat Map with continuous variables



Inference:

```
loan_amount, funded_amount,
funded_amount_inv is highly correrlated
#funded_amount has negative correlation
with loan Status
#intrest Rate is also having  negative
correlation with loan Status
#Monthly_Installment_IncomeRatio is also
having  negative correlation with loan
Status
```

Recommendations:

Following attrubutes looks to be important:
funded_amount, verification_status,
employment_length, grade

1. More Monthly_Installment_IncomeRatio, means chances of being Charged Off increases
2. Employment length has inverse relation with loan status.. so lesser the employment length, higher is the chances of getting converted to a bad loan.
3. If the grade given to borrower is A, B, C then chances of loan being charged off decreases.
4. Loans with Source of income verified has lesser chances of being charged off