Lending Club
Case Study

Group Members:
Soumita Das & Jafar Naqvi

Table of Contents

1 Problem Statement and Analysis approach

2 Univariate Analysis

3 Segmented Univariate Analysis

4 Bivariate Analysis

5 Multivariate Analysis/Correlation

6 Conclusions

1

Problem Statement and Analysis approach

Problem Statement

Lending money involves inherent risks, and when borrowers fail to repay, it results in significant financial losses for the lending company, ultimately impacting the company's cash flow. The ability to identify these high-risk applicants is crucial for minimizing credit losses and preserving the financial health of the organization

1 Problem Statement and Analysis approach

Analysis Approach

Step 1 - Data Cleaning

Step 2 - Data Transformation

Step 3 - Univariate Analysis

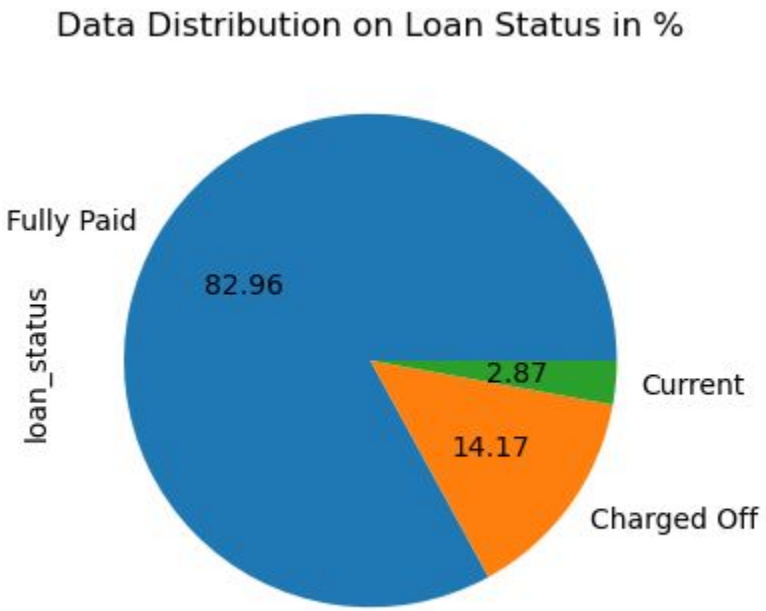
Step 4 - Segmented Univariate Analysis

Step 5 - Bivariate Analysis

Step 6 - Multivariate Analysis

Step 7 - Conclusion

1 Problem Statement and Analysis approach

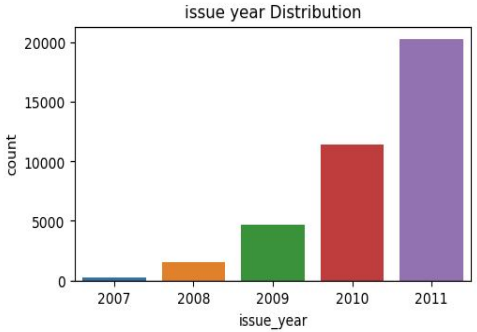


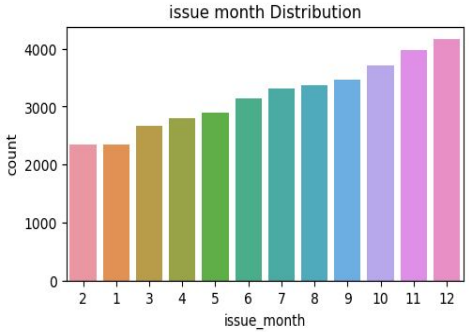
The percentage wise distribution of loans:

83% - Fully Paid
14% - Charged Off (Defaulters)
3% - Current

For Analysis we will remove Current Loan status and proceed with the the Charged off and Fully Paid loans

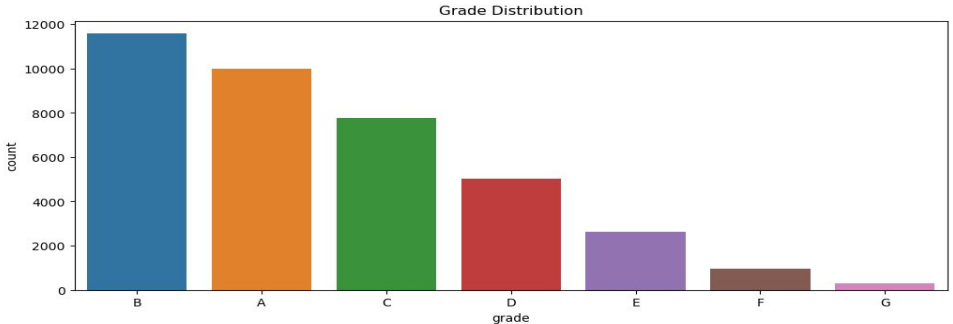
2 Univariate Analysis





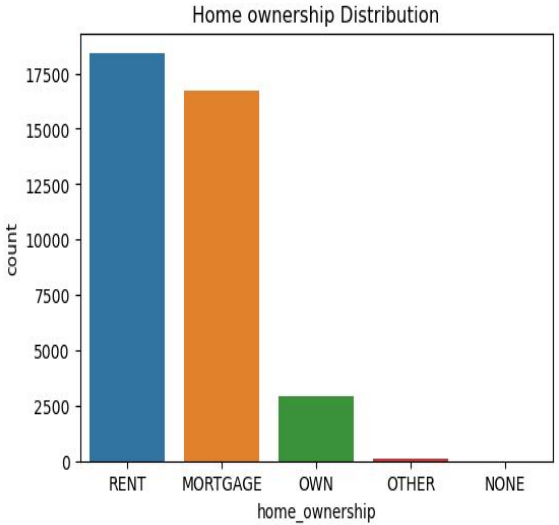
-From the issue year distribution, the loans have increased year on year almost double

-From the issue month distribution, most loans were borrowed in last couple of months (Nov,Dec) , this might by due to festival holiday season.

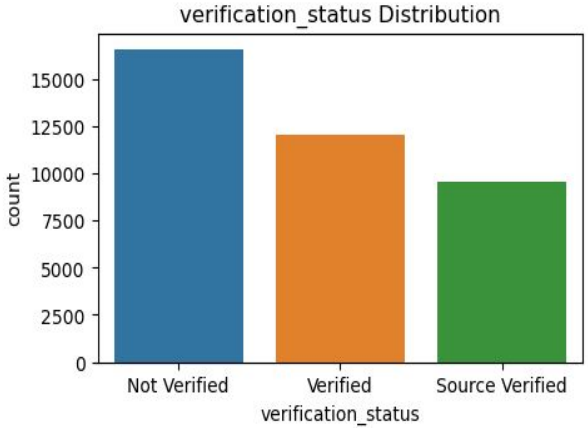


From the Grade distribution, most loans were applied under grade A and B

2 Univariate Analysis

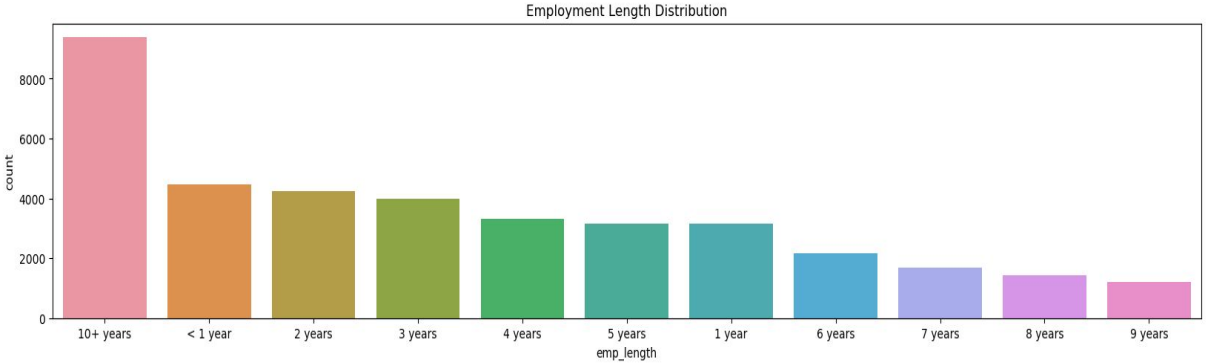


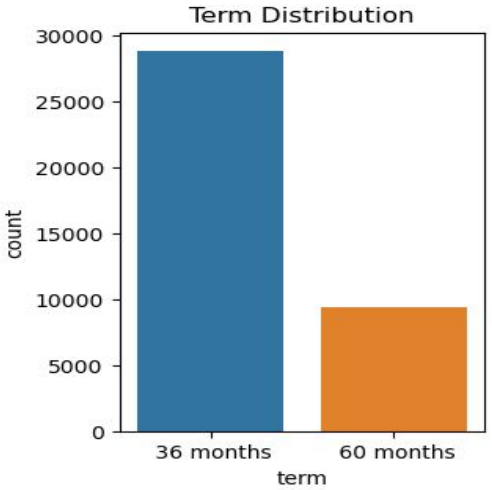
-From home ownership distribution, most loans were applied where Borrower have Rented home, followed by borrowers who have mortgage.



-From the verification status distribution, most loans were given without borrower's income verification
-If we add Verified (verified by LC) and Source Verified cases then overall Verified is more than Non Verified cases.

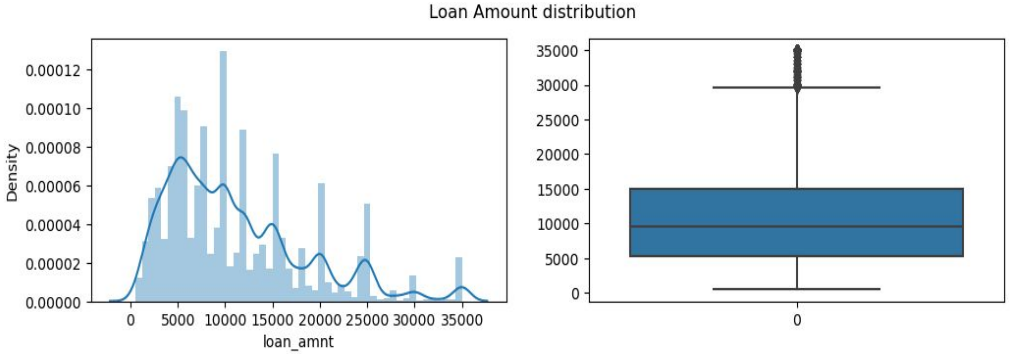
2 Univariate Analysis





- The employment length for most borrowers is more than 10 years
- Number of borrowers increases with the increase in employment
- Most people took loan for 36 months duration

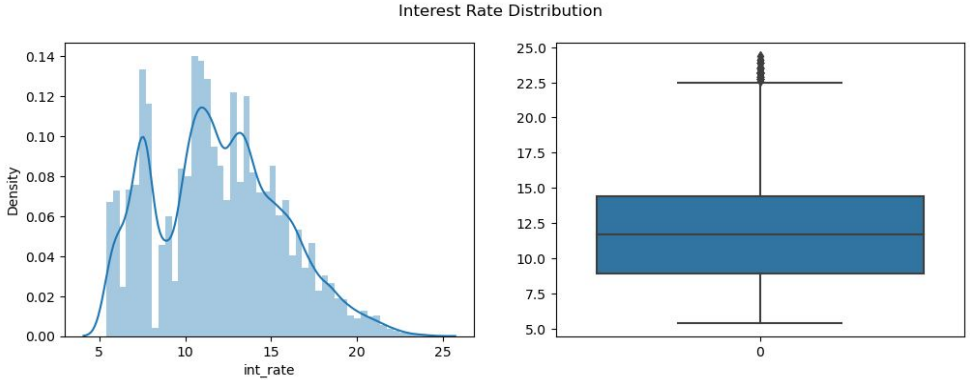
2 Univariate Analysis



-From Loan amount plots, most people took loan amount of 10000, and median of distribution is 10000.

-Very few people took more than 30000 loan amount.

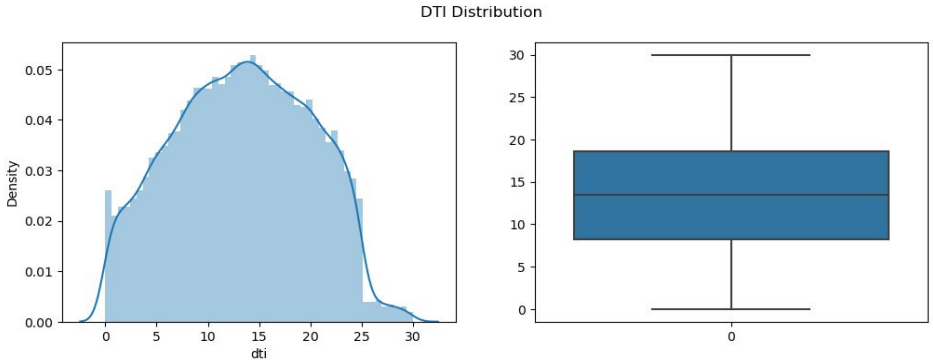
-There seems to be no outlier since median and mean does not have very big difference and this difference is less than standard deviation.



- From interest rate plots, most people took loan with the interest rate ranging between 8.9% to 14.4 %

- There are few people who have taken loans on higher interest rates of 22.5%

2 Univariate Analysis

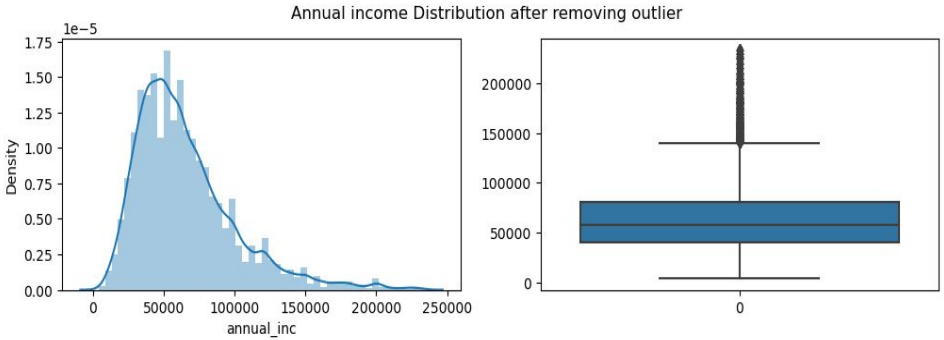


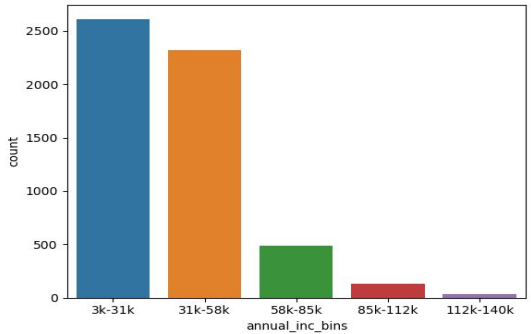
-The DTI distribution is similar to normal distribution

-There are no outliers for DTI

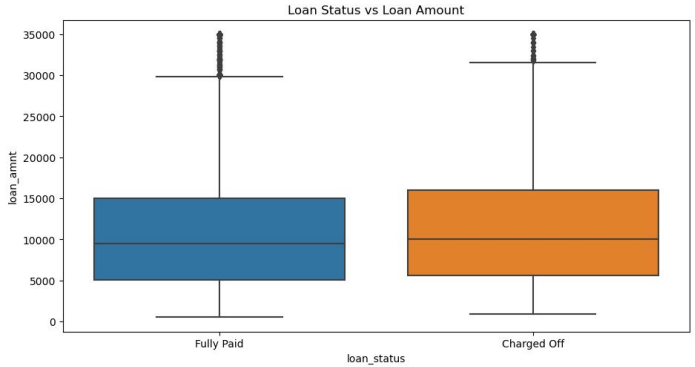
- The distribution shows that most borrowers has debt to income ratio within 30

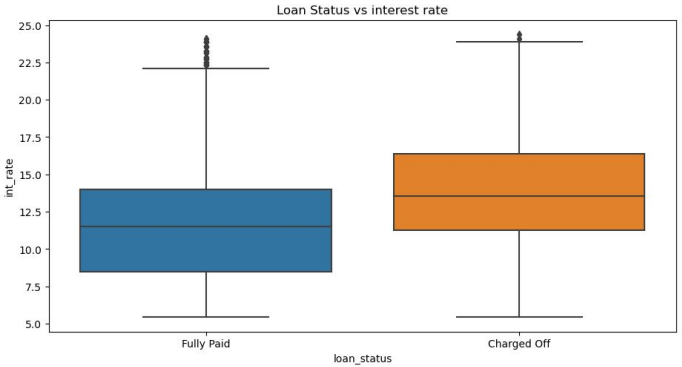
- Most people took loan whose salary is upto 30k

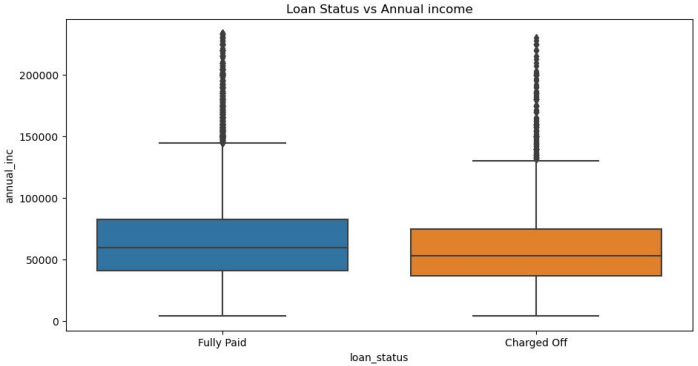


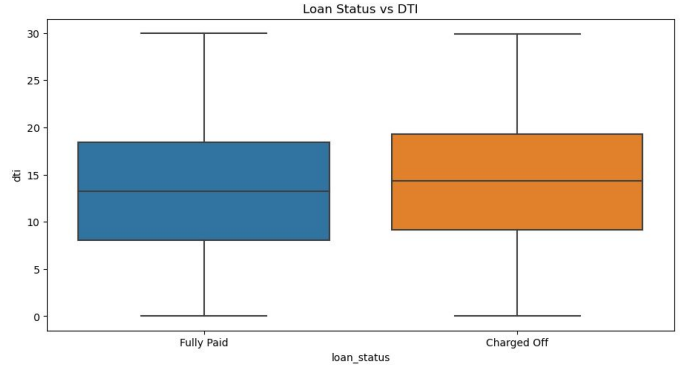


3 Segmented Univariate Analysis









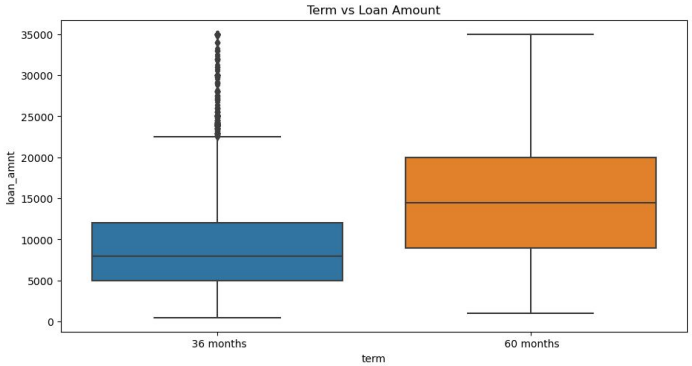
-Charged Off loans have higher loan amounts than Fully Paid ones

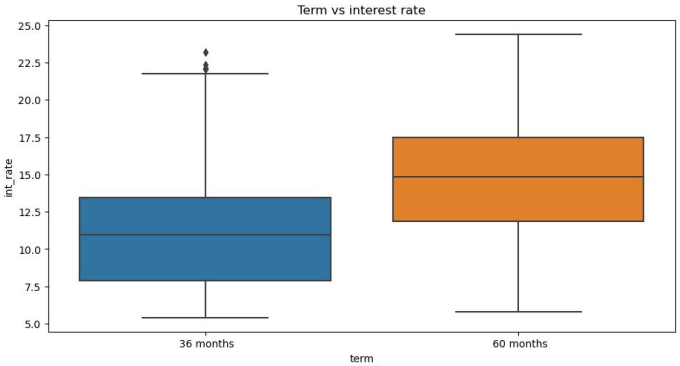
-Fully Paid loans have higher annual income than charged off ones

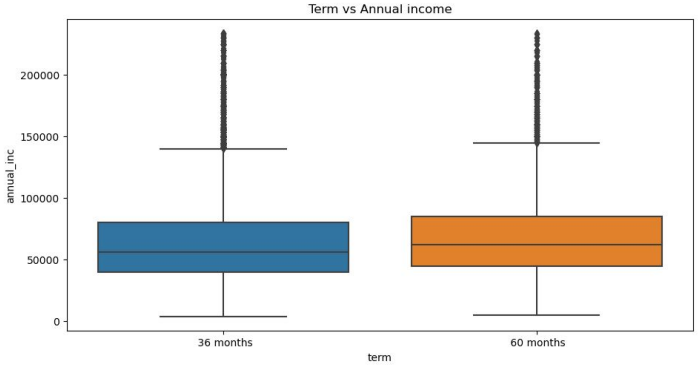
-Interest rate for charged off loans is higher than Fully paid loans

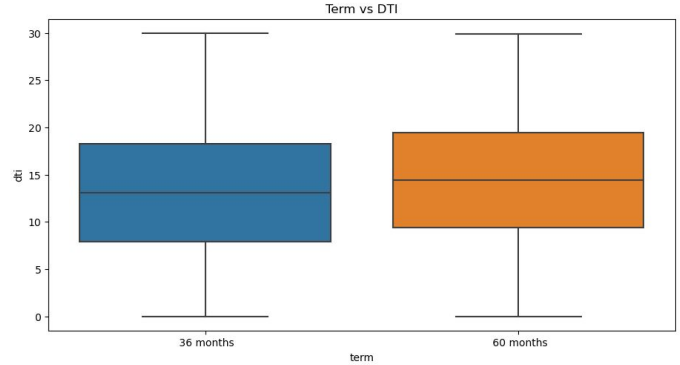
-DTI is higher for charged off than Fully paid ones

3 Segmented Univariate Analysis









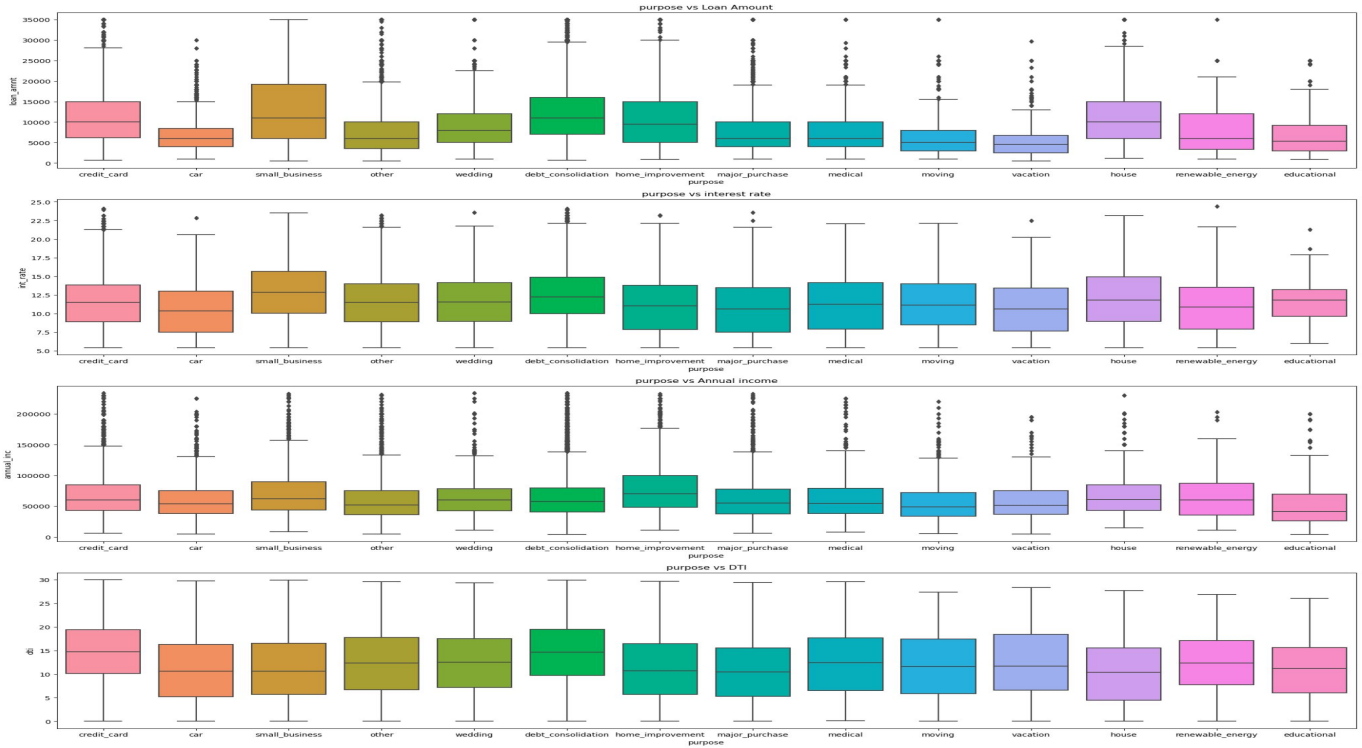
-Loan Amount for 60 months Term is higher than 36 months

-Loan Interest rate for 60 months term is higher than 36 months

-Annual income of borrowers is slightly higher for 60 months loan as compared to 36 months

-DTI for 60 months term loan is higher than 36 months term

3 Segmented Univariate Analysis

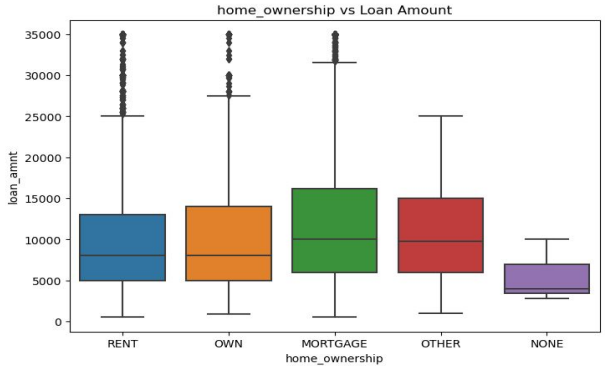


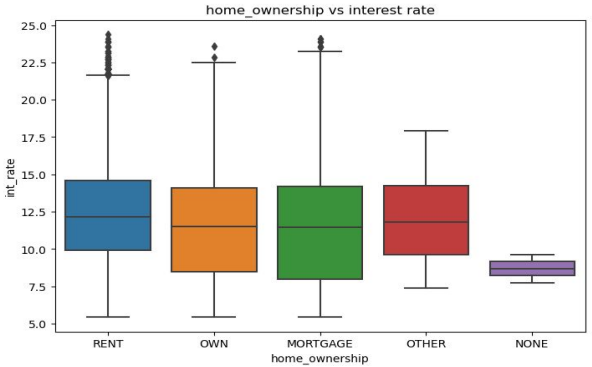
-Small businesses were given highest mean loan amount, also various range of loan amount were given to Small businesses

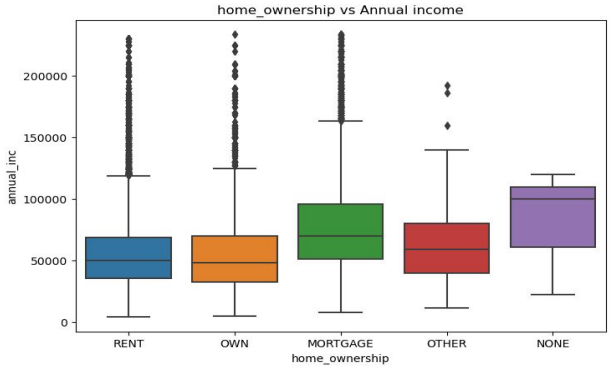
-Loan purpose of Small businesses and Debt consolidation were offered highest interest rates

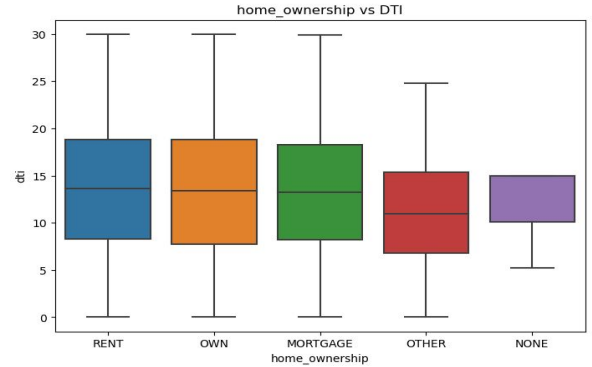
-Annual income for borrowers is highest for loan purpose of home improvements-DTI is highest for debt_consolidation

3 Segmented Univariate Analysis









-highest loan amount was offered to borrowers with Mortgage

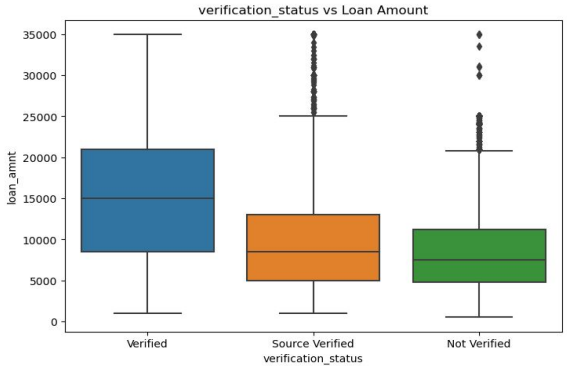
-highest interest rate is for rented borrowers

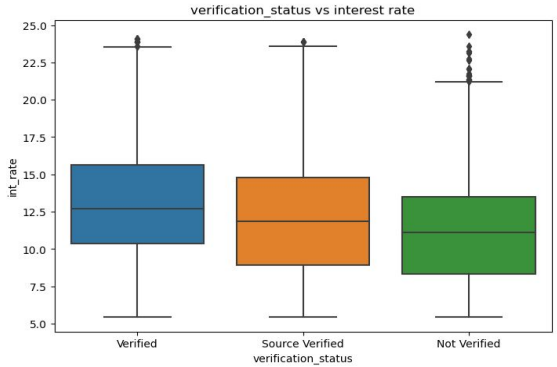
-Borrowers with mortgage were offered varied interest rates ranges

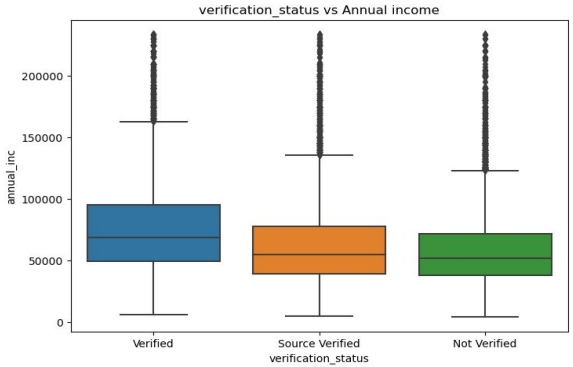
-DTI is almost same across borrowers with Rent,Own and Mortgage

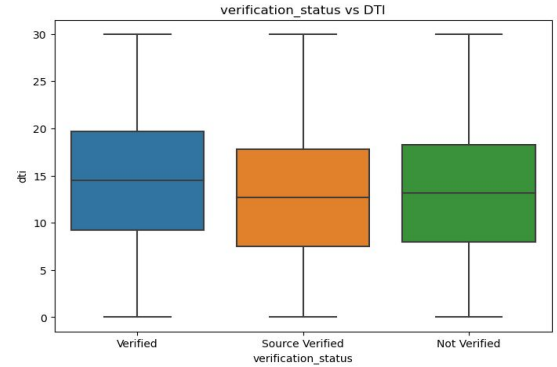
3

Segmented Univariate Analysis









-Lowest loan amount was lended for Not verified borrowers.Highest amount were given to Verified borrowers.

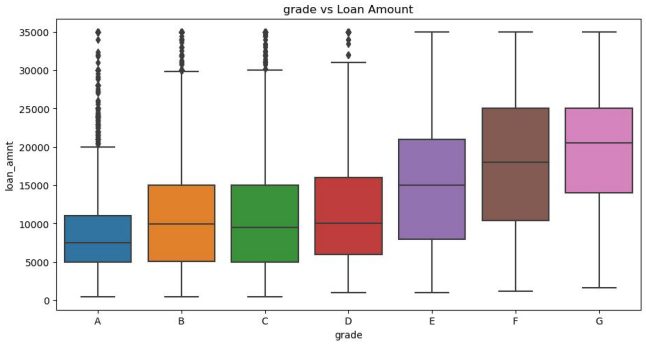
-interest rate is lowest for not verified cases as compared to verified cases

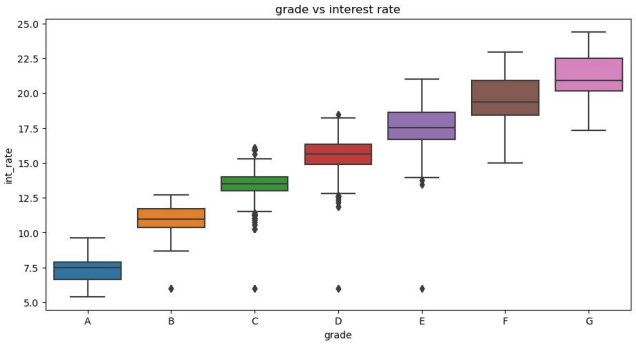
-DTI is highest for Verified cases

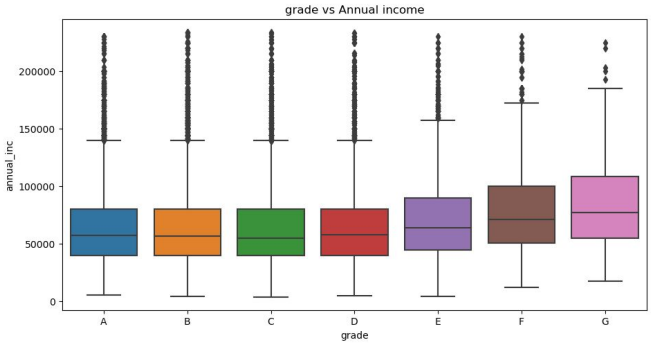
-annual income is highest for Verified cases

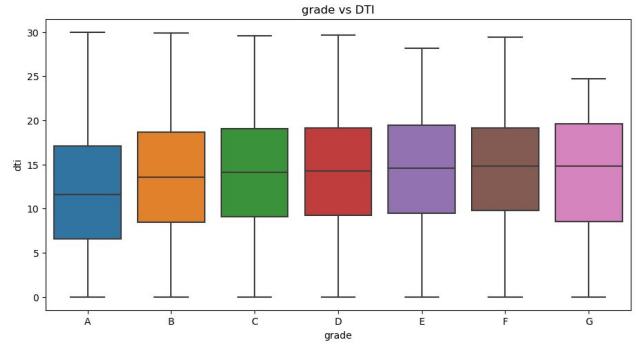
3

Segmented Univariate Analysis









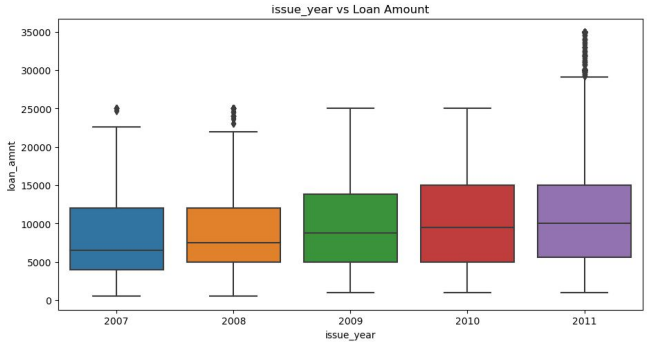
-Loan amount is highest for grade G followed by grade F. As the Grade decreases the loan amount increased

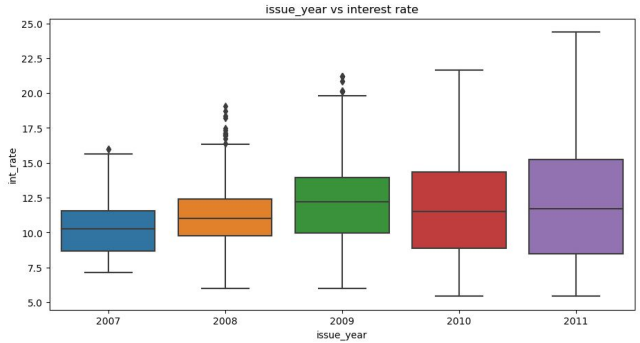
-Interest rate increases from Grade A to G. Diverse interest rates present for lower grades F,G

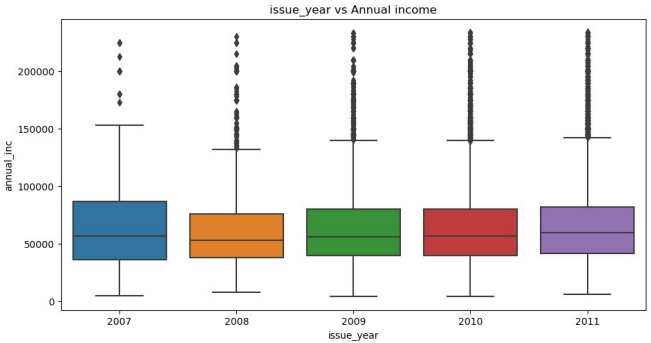
-lower grades have diverse annual incomes

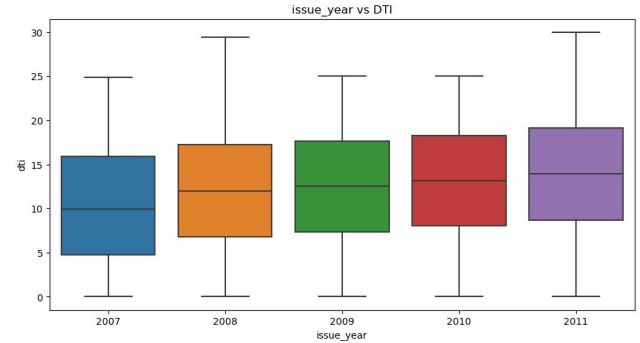
- DTI is least for Grade A borrowers

3 Segmented Univariate Analysis









-loan amount have increased gradually over the years from 2007 to 2011 since the mean is increasing year on year

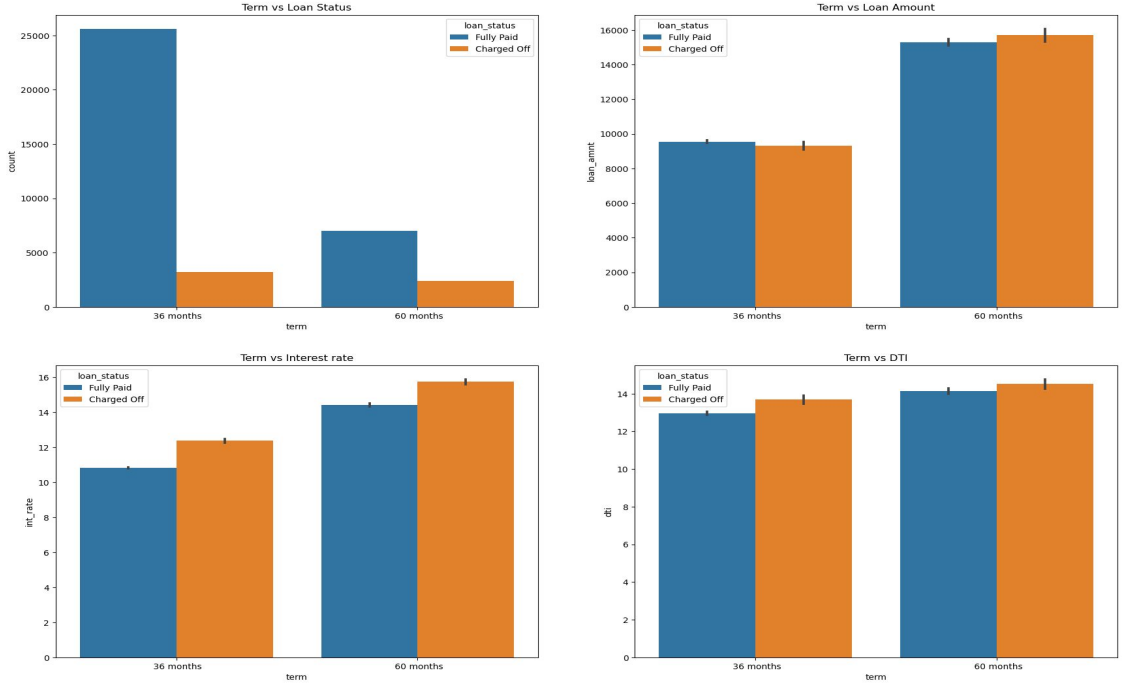
-DTI have gradually increased over the years

-Interest rates are diverse over the years

-Annual income almost same across the years

4 Bivariate Analysis

Term Vs Categorical & Numerical Variables



-Fully Paid is higher in 36 months term

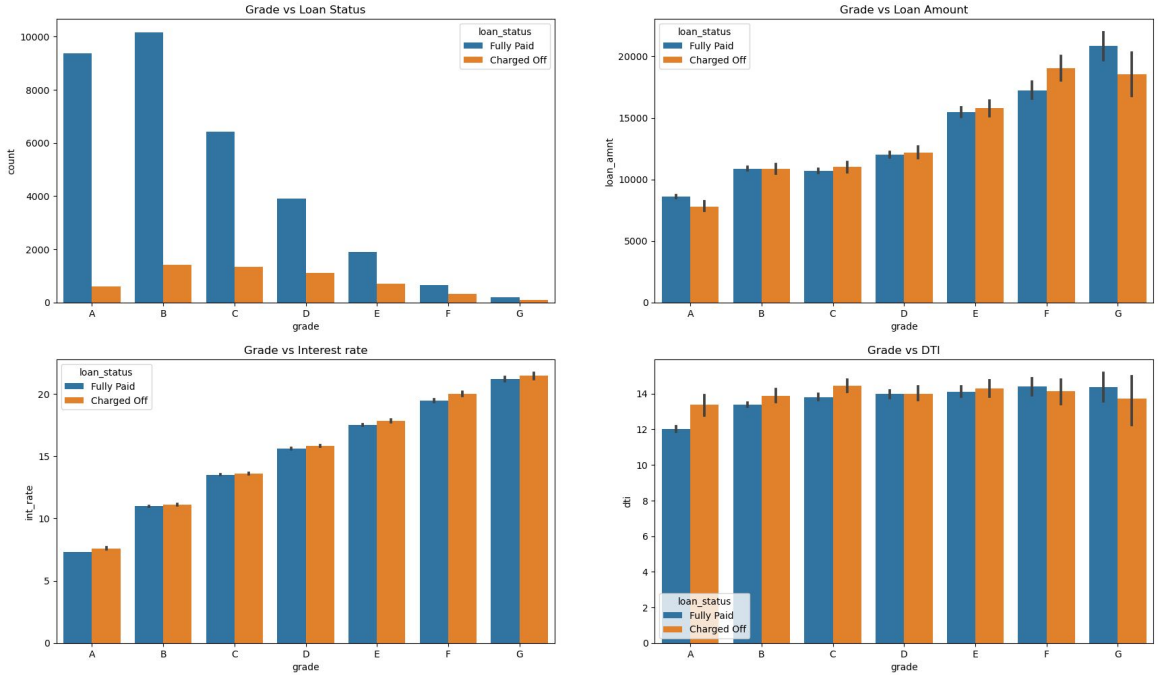
-Loan amount does not serve as a determining factor for defaults in both the 36-month and 60-month loan tenures. Borrowers exhibit a balanced distribution between defaults and non-defaults across both of these tenures.

-When interest rates are elevated, the default rate increases for both the 36-month and 60-month loan tenures.

-charge-off are higher when compared with fully-paid for the Debt to income ratio

4 Bivariate Analysis

Grade Vs Categorical & Numerical Variables



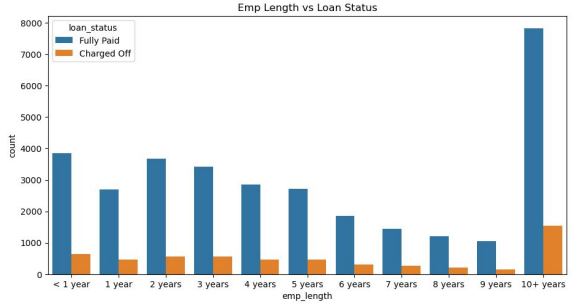
-As grade decreases Charge Offs interest rates increates

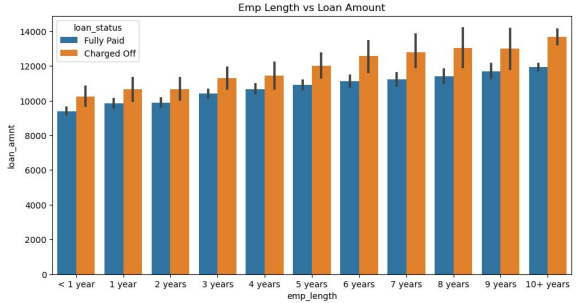
-As grade decreases loan amount increases who are more prone to default

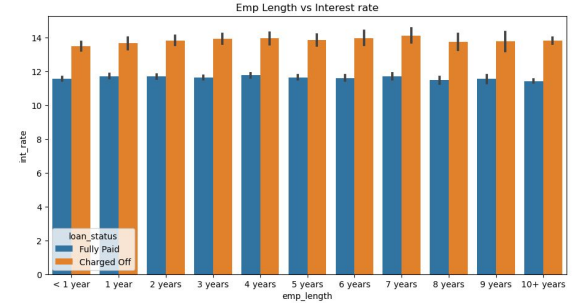
-For grade F and G the difference between Fully paid and Charge offs is higher

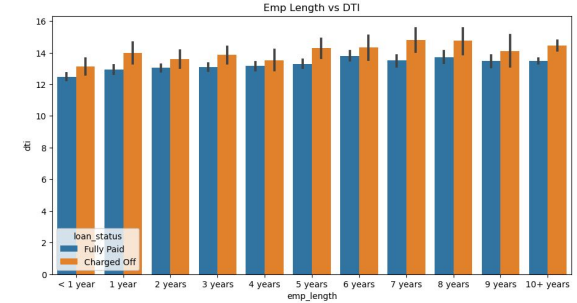
4 Bivariate Analysis

Emp Length Vs Categorical & Numerical Variables









-No pattern observed for defaulters across employment lengths

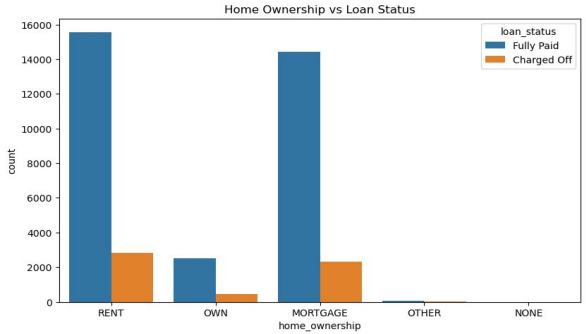
-Borrowers with higher employment lengths have taken high loan amounts and also got more default rate.

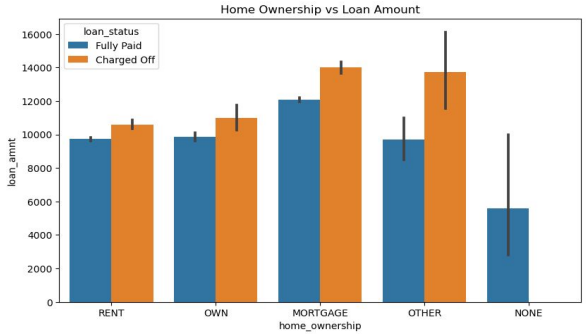
-The difference of interest rates between Fully Paid Vs Defaulters is consistent across employment length

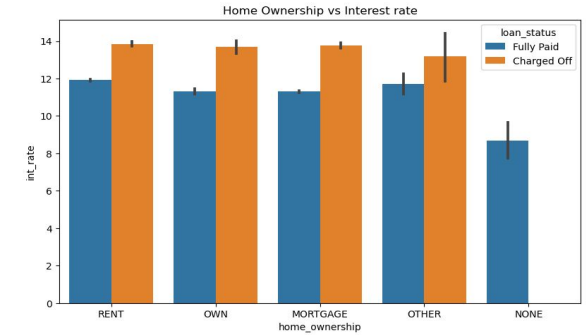
-no patterns observed for DTI

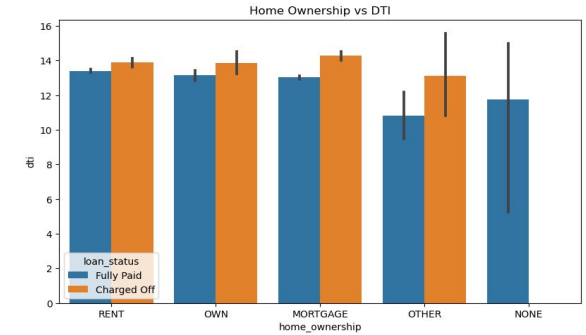
4 Bivariate Analysis

Home Ownership Vs Categorical & Numerical Variables







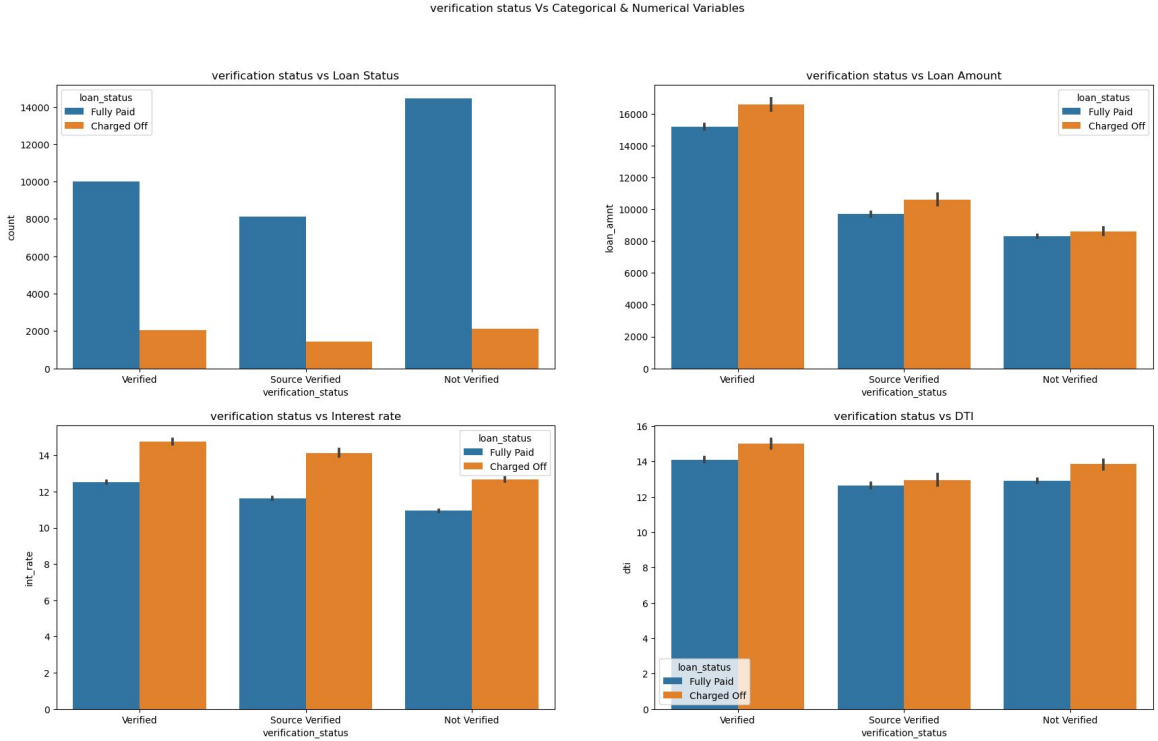


-Rent category have higher charged off as compared to other categories

-Borrowers with Mortgage have higher defaulters as compared to other categories

-Interest rate is higher for rented borrowers which tends to default more

4 Bivariate Analysis



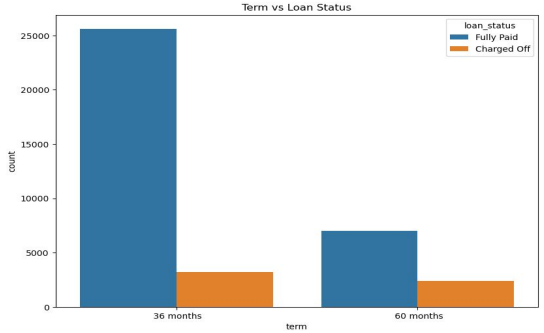
-Verified were given more loan amounts, also Defaulters are more in verified category

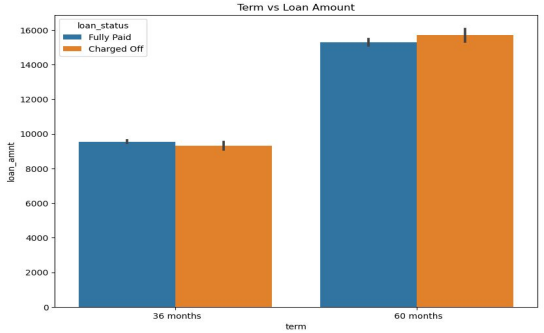
-Interest offered for Verified is comparatively more

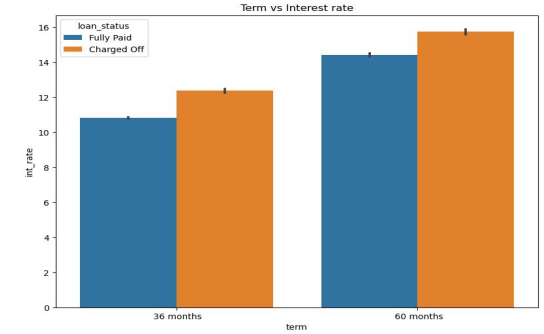
-DTI is higher for Verified category

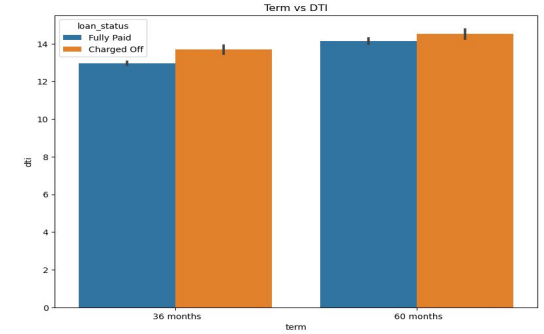
4 Bivariate Analysis

Term Vs Categorical & Numerical Variables









-Fully Paid is higher in 36 months term

-Loan amount does not serve as a determining factor for defaults in both the 36-month and 60-month loan tenures. Borrowers exhibit a balanced distribution between defaults and non-defaults across both of these tenures.

-When interest rates are elevated, the default rate increases for both the 36-month and 60-month loan tenures.

-Charged-off are higher when compared with fully-paid for the Debt to income ratio

4

Bivariate Analysis

Purpose Vs Categorical & Numerical Variables

-More loans were borrowed for Debt consolidation

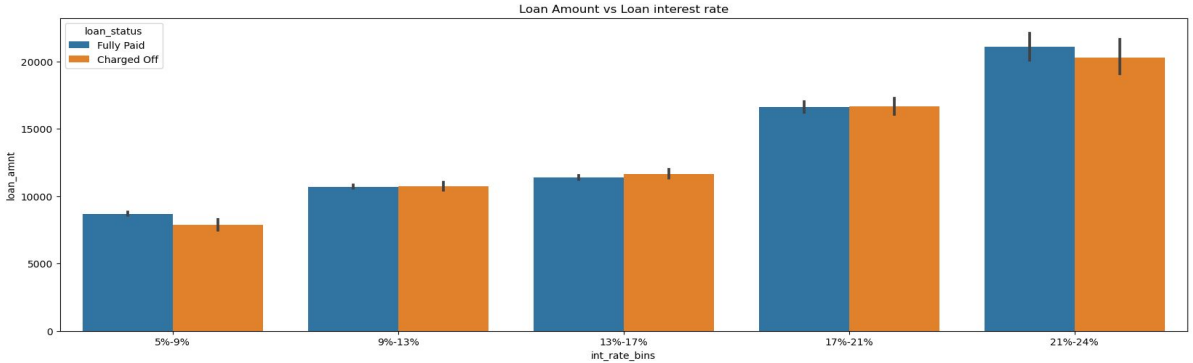
-More defaulters have defaulted in Debt consolidation loan purpose

-Interest rate for House is higher among defaulters followed by debt consolidation

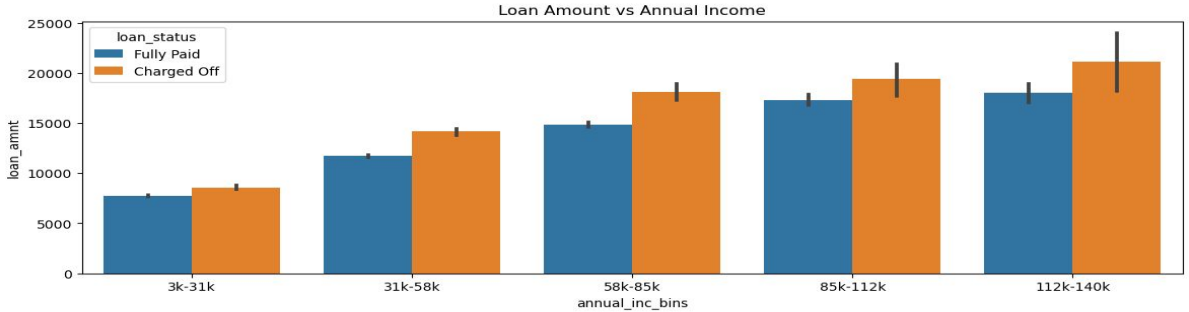
-DTI is higher for purpose - credit card and debt consolidation

-Loan amount is higher for defaulters against fully paid in all categories except moving category

4 Bivariate Analysis



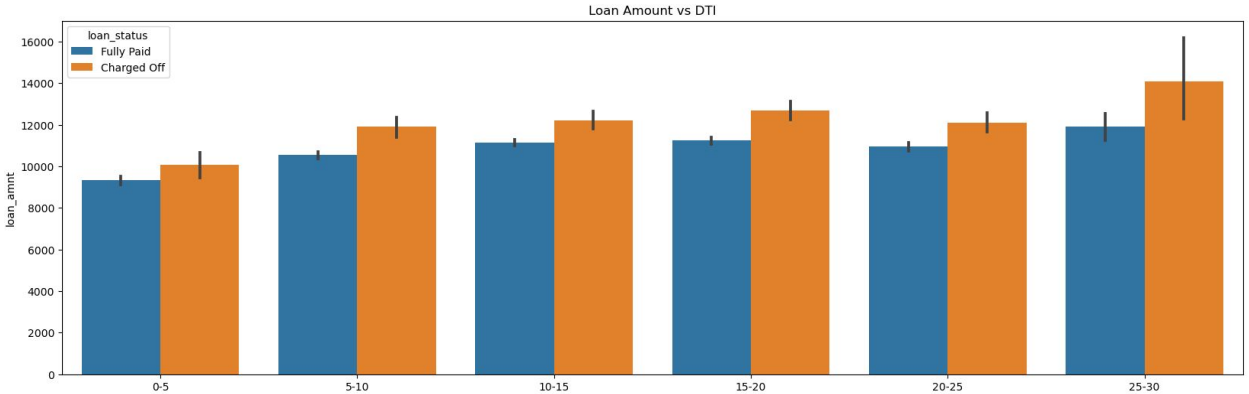
With increase in loan amount the interest rates increased i.e. higher loans were given for higher rates



-With increase in annual income the loan amount increases.

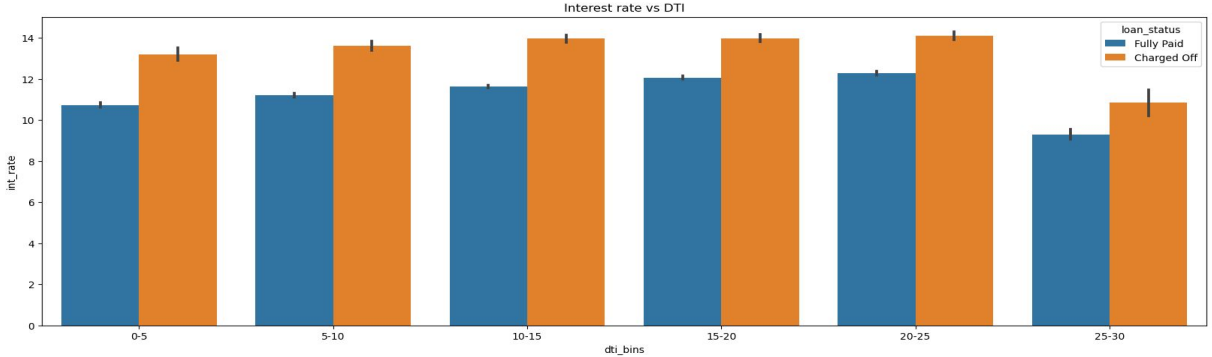
-The defaulters are highest for the highest income groups.

4 Bivariate Analysis



-With increase in DTI the loan amount increases.

-The defaulters are highest for the highest DTI group (25-30).

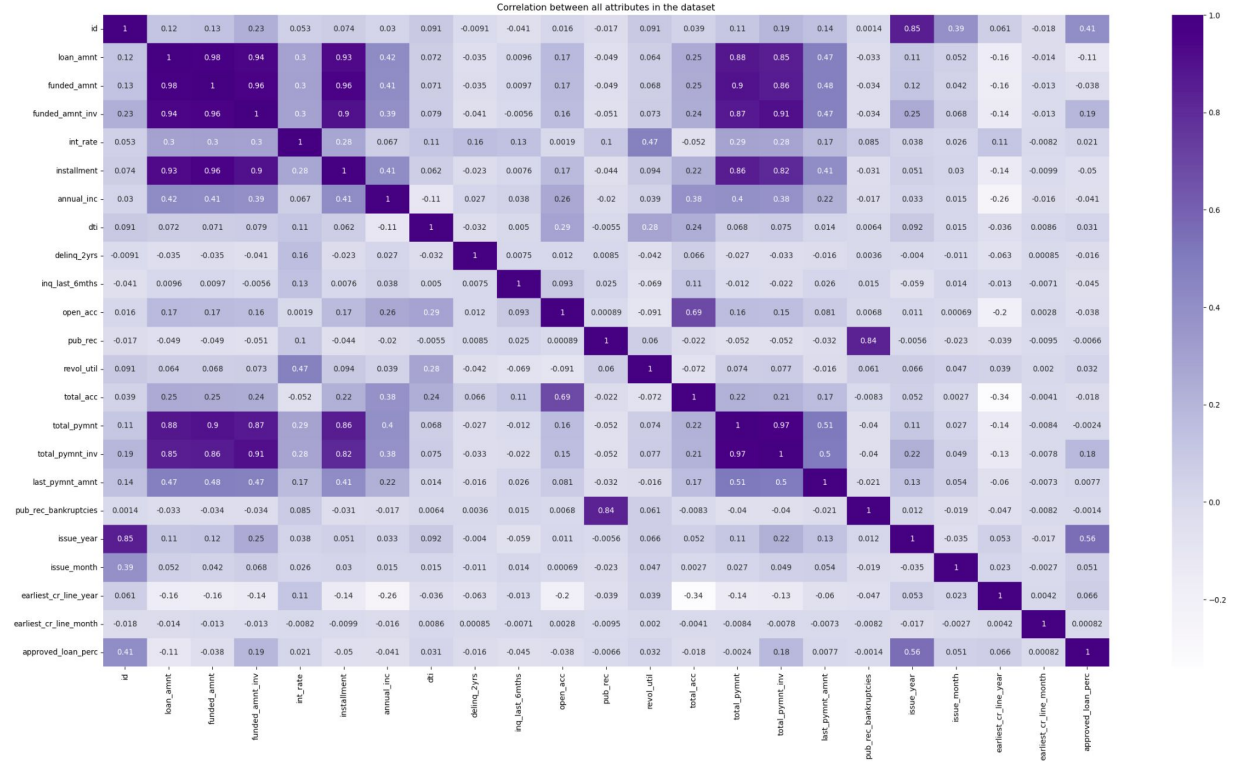


-Interest rate is lowest for highest DTI group

-Defaulters are lowest for highest DTI group

5

Multivariate Analysis/Correlation Matrix



- loan_amnt is highly positively correlated with funded_amnt_inv,installment, total_pymnt_inv.
- loan_amnt has intermediate positive correlation with annual_inc, int_rate, total_acc.
- pub_rec is highly positively correlated with pub_rec_bankruptcies.

6 Conclusions

- Small business loans experience a higher default rate, suggesting that Lending Club should consider discontinuing or reducing the issuance of such loans.
- Grades serve as effective metrics for identifying potential defaulters, prompting Lending Club to consider gathering more comprehensive borrower information before extending loans to those with lower grades (G to A)
- Enforce a loan cap of 12,000 for applicants who possess mortgage ownership. This measure also serves as an indicator of the financial stability of high-risk applicants.
- Apply an additional interest rate to the loan amount for applicants who state their purpose as 'moving,' particularly for smaller loan amounts, as they are more likely to demonstrate a higher repayment likelihood
- Lending Club should consider lowering the interest rates on 60-month tenure loans, as they exhibit a higher risk of loan defaults.
- loan issues to borrowers who are from CA, FL and NY should be restricted since they have higher default rates