

Lending Club Case Study



Subhomoy Ganguli
Swati Gupta Jain

Email

subhomoyganguli1001@gmail.com

swatigupta0106@gmail.com

Problem Statement

Lending Club 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.

Using Exploratory Data Analysis (EDA), Lending Club wants to understand the driving factors (or driver variables) behind loan default, i.e., the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.

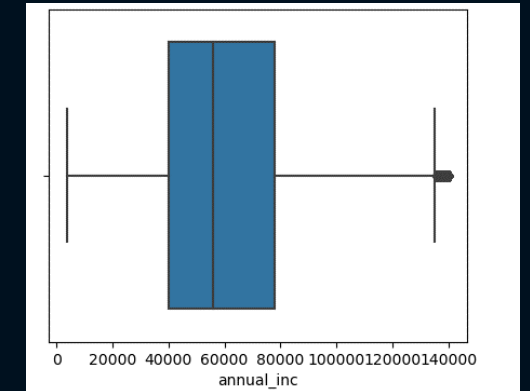
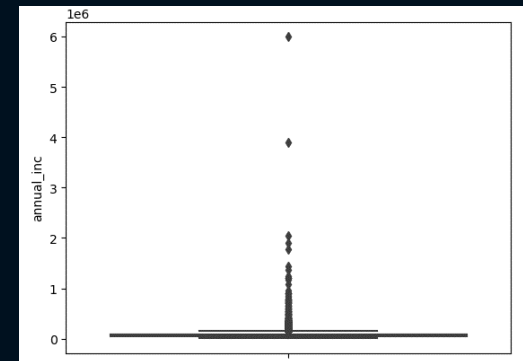
The EDA will help to understand how consumer attributes and loan attributes influence the tendency of default.



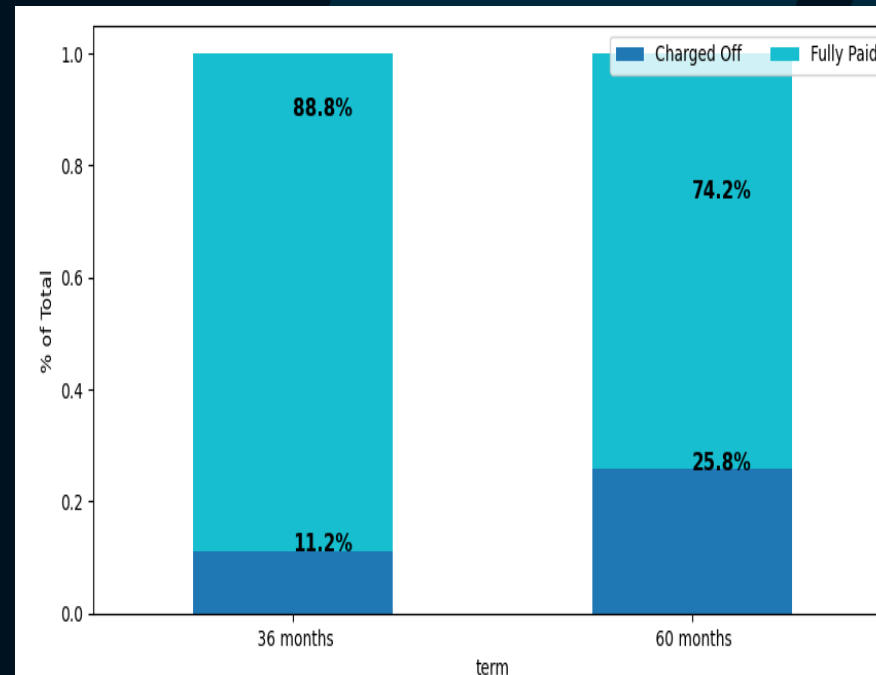
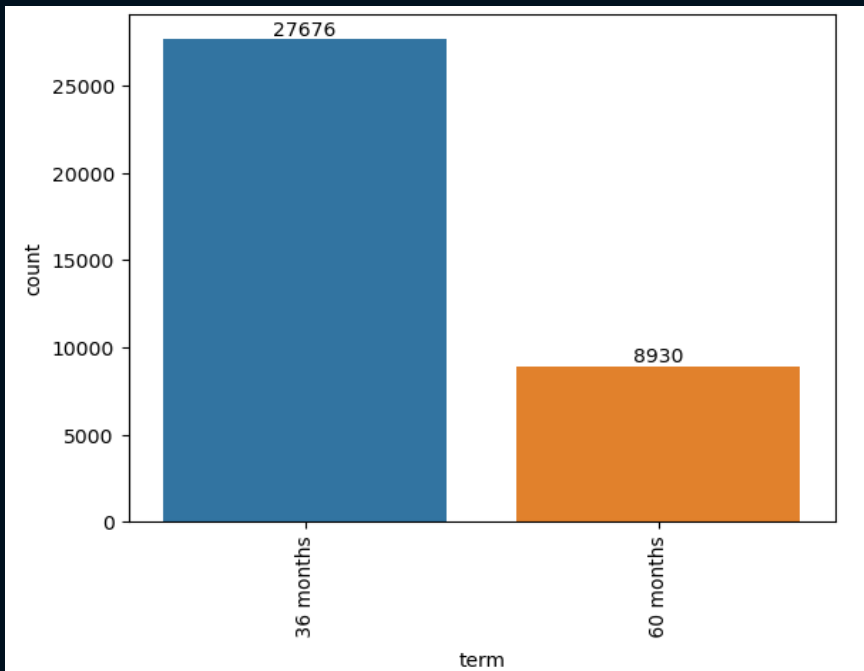
Data Cleaning

- Dropping columns which have only null values or are single valued or have more than 60% of null
- Dropping columns which are unique in nature like member_id, url, etc. as these don't help in the analysis
- Standardizing the data so that integer, float or date columns are represented properly
- Handling outliers, using box plot, we identified outlier in some dimensions and using quartile we cleaned these outliers

Example of Outlier Treatment



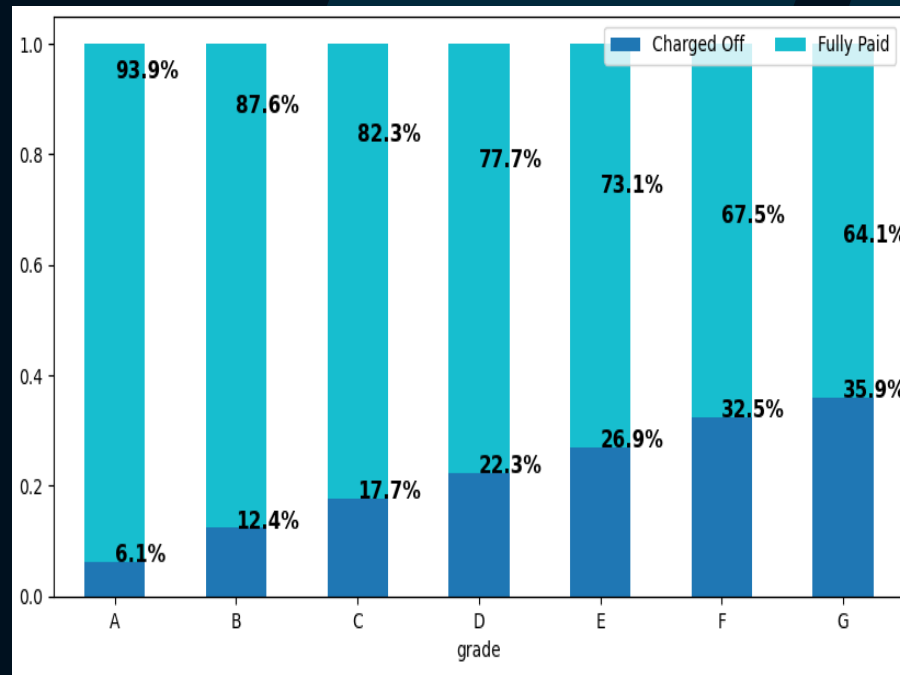
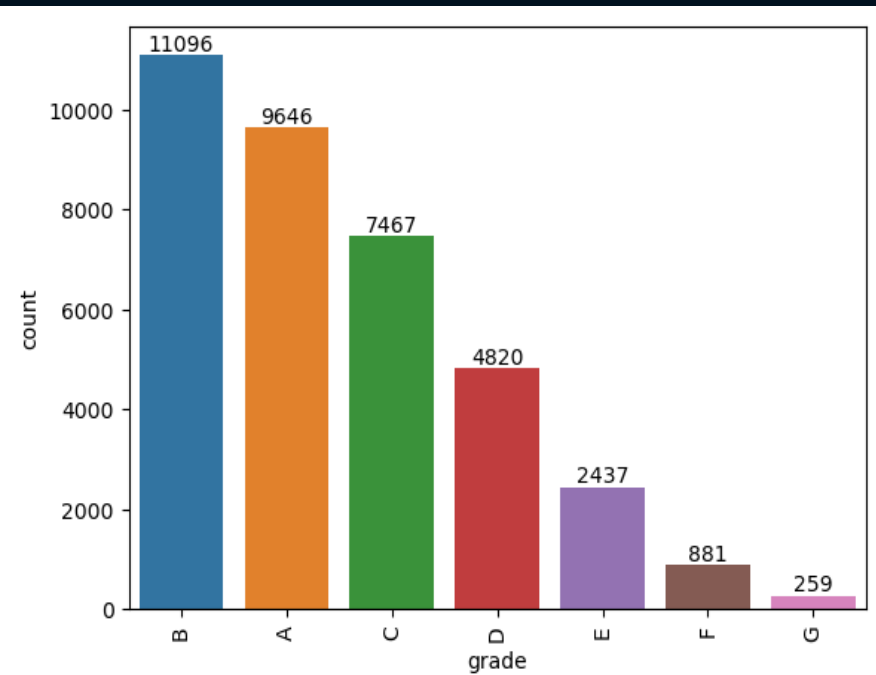
Analysis - Term



Observations

- 36 months term loans are more common than 60 months
- 60 months term loans are more likely to default

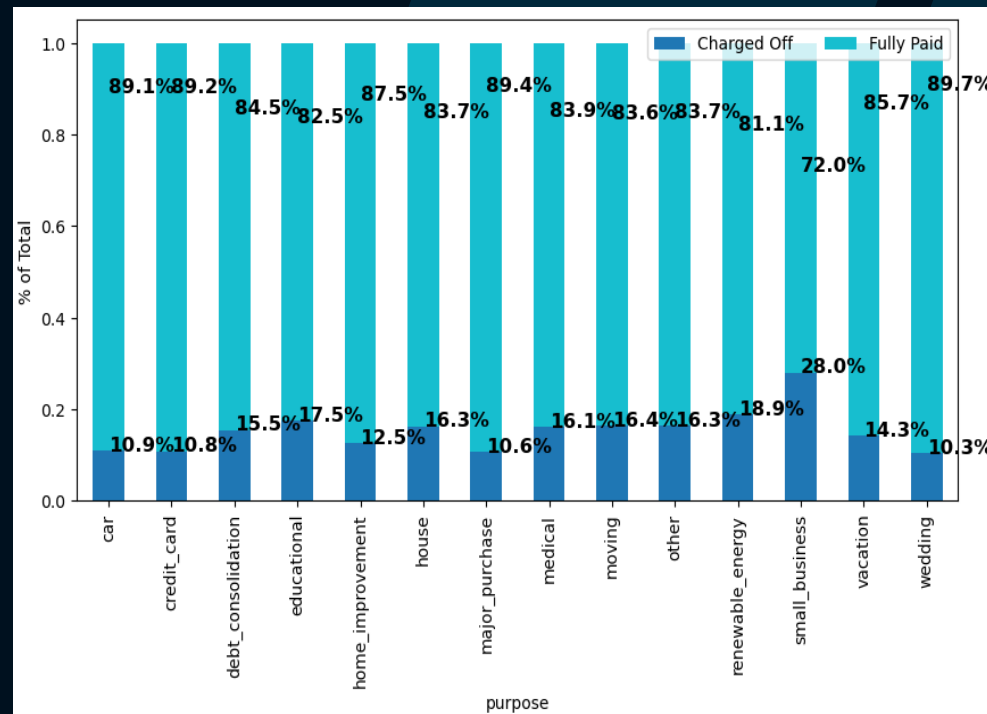
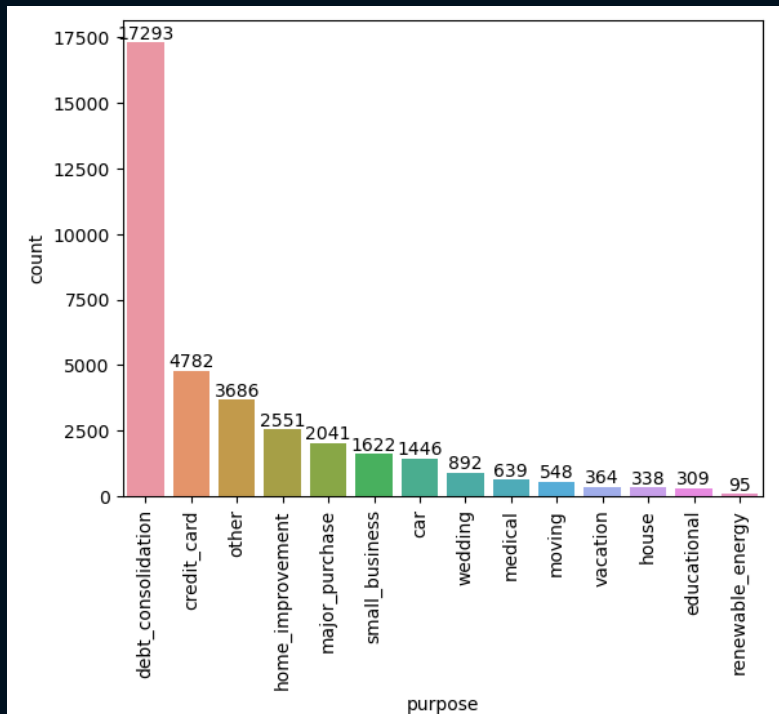
Analysis - Grade



Observations

- More than 50% of loans are of high grade, A or B
- Grade and chances of default are inversely proportional. We can say Lending Club is quite good at grading its loans

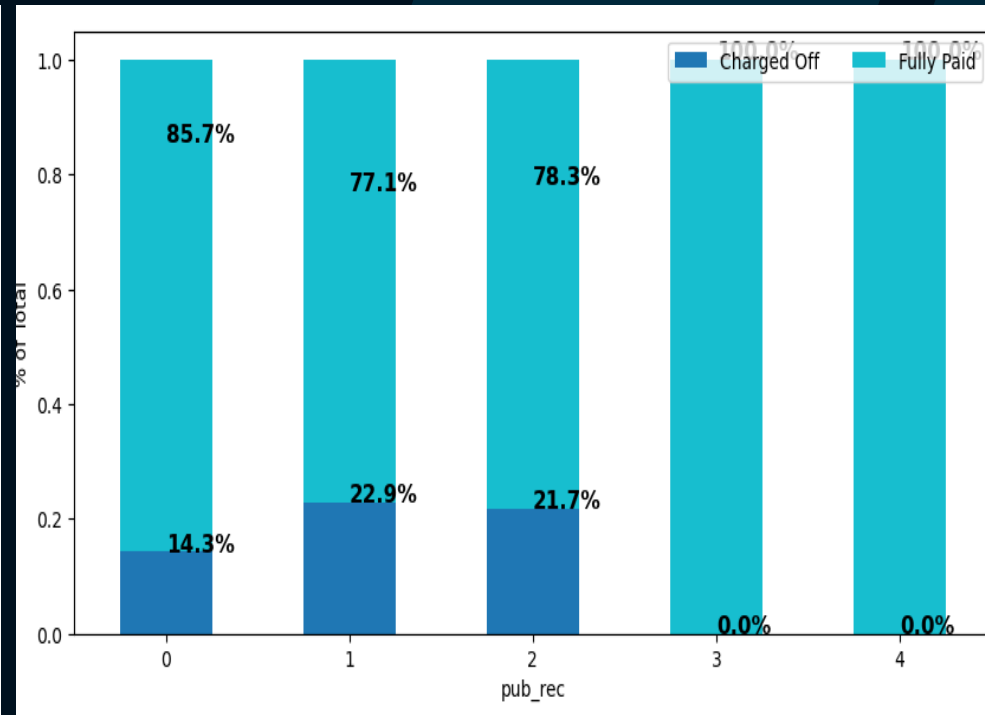
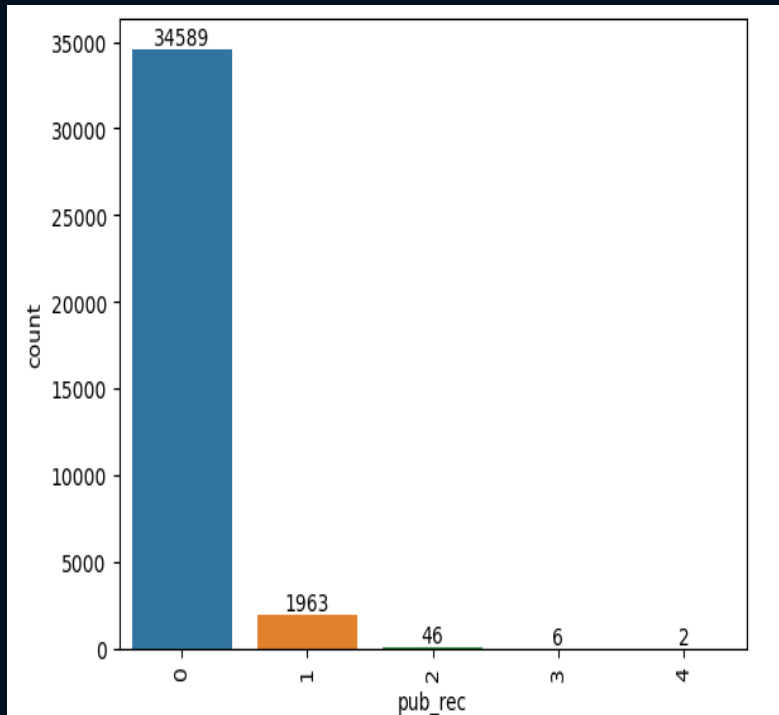
Analysis - Purpose



Observations

- Debt consolidation comprises of almost 49% of the loans, followed by credit card
- 28% of loans to small businesses are defaulted, making them the riskiest ones

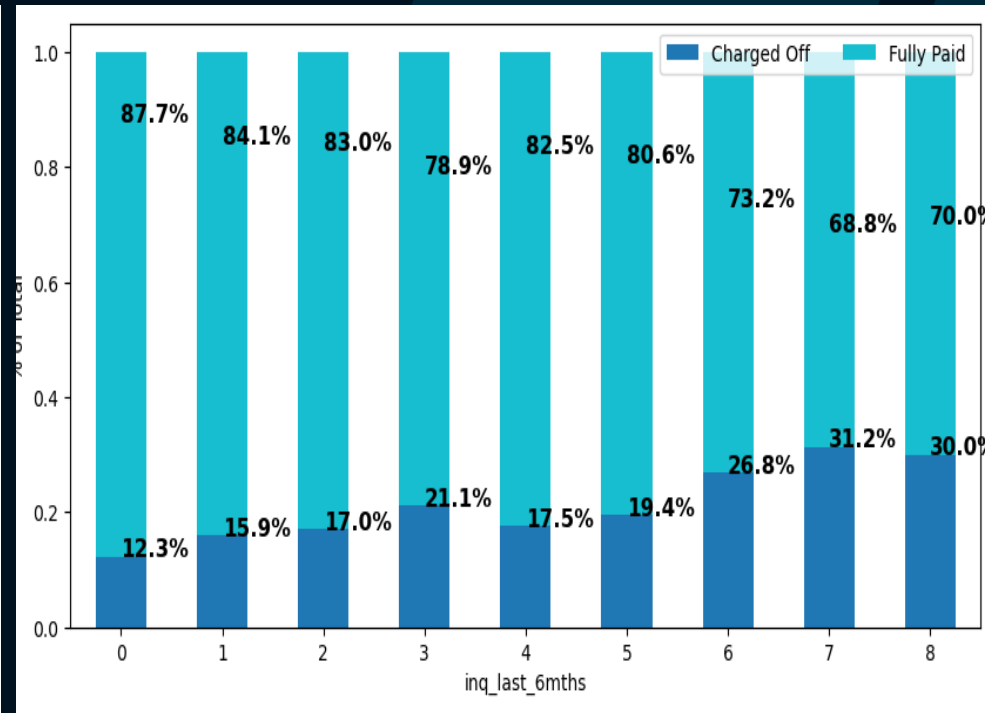
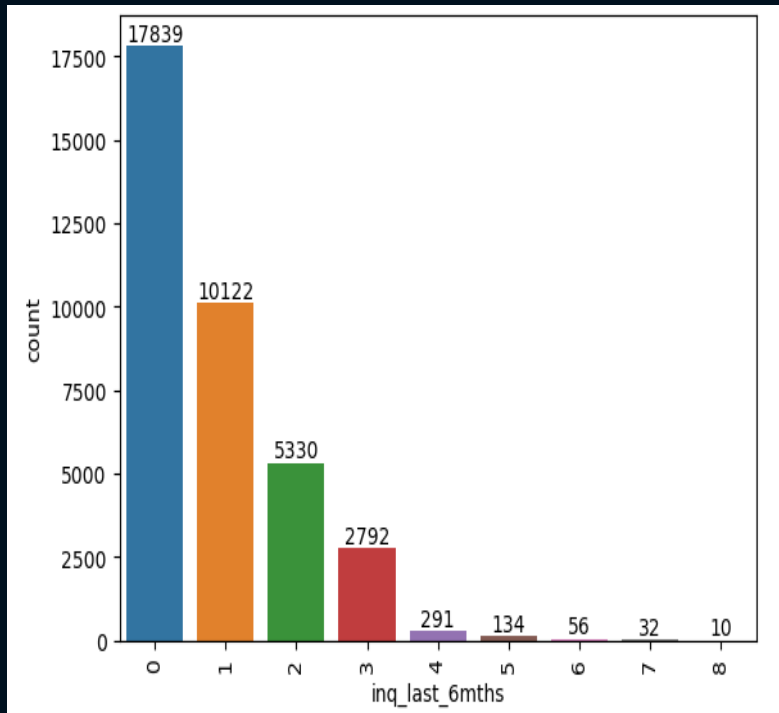
Analysis – Public Record



Observations

- Chances of default increases significantly when borrower has prior derogatory public record

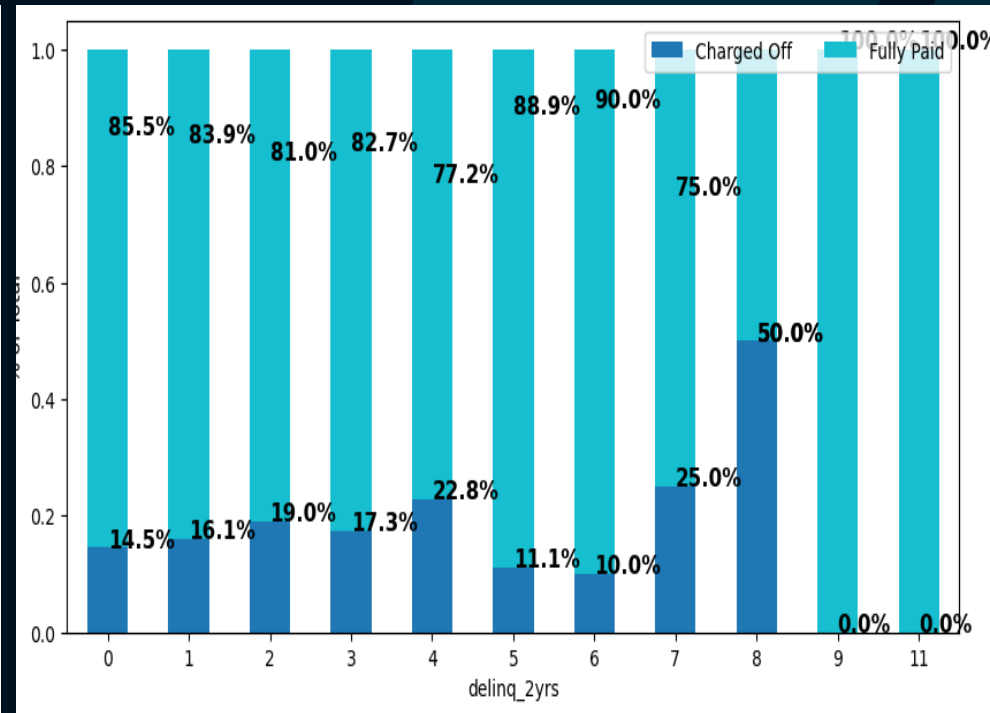
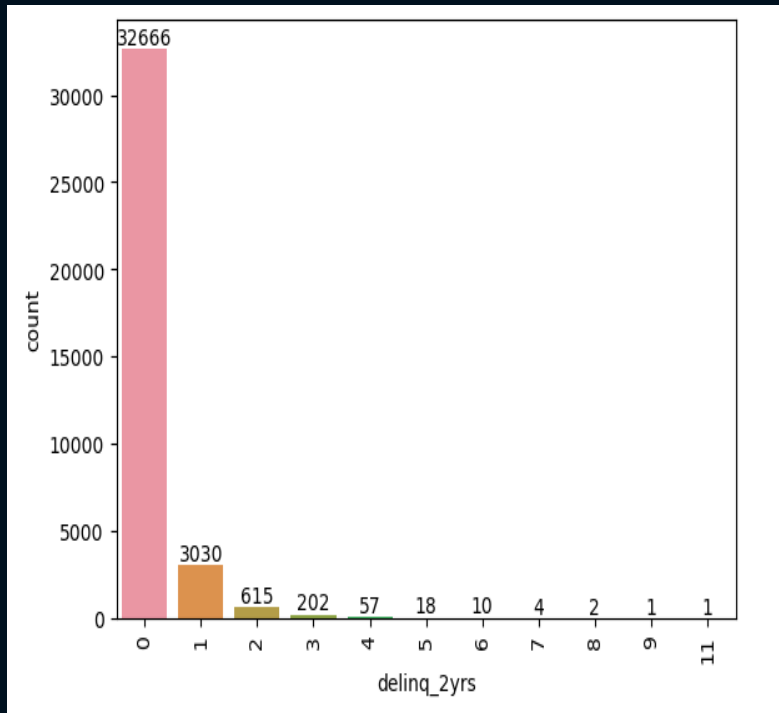
Analysis – Inquiry in last 6 months



Observations

- Most of the loans never had any inquiry and are less likely to default
- Loans with more than 2 inquiries have higher rate of default

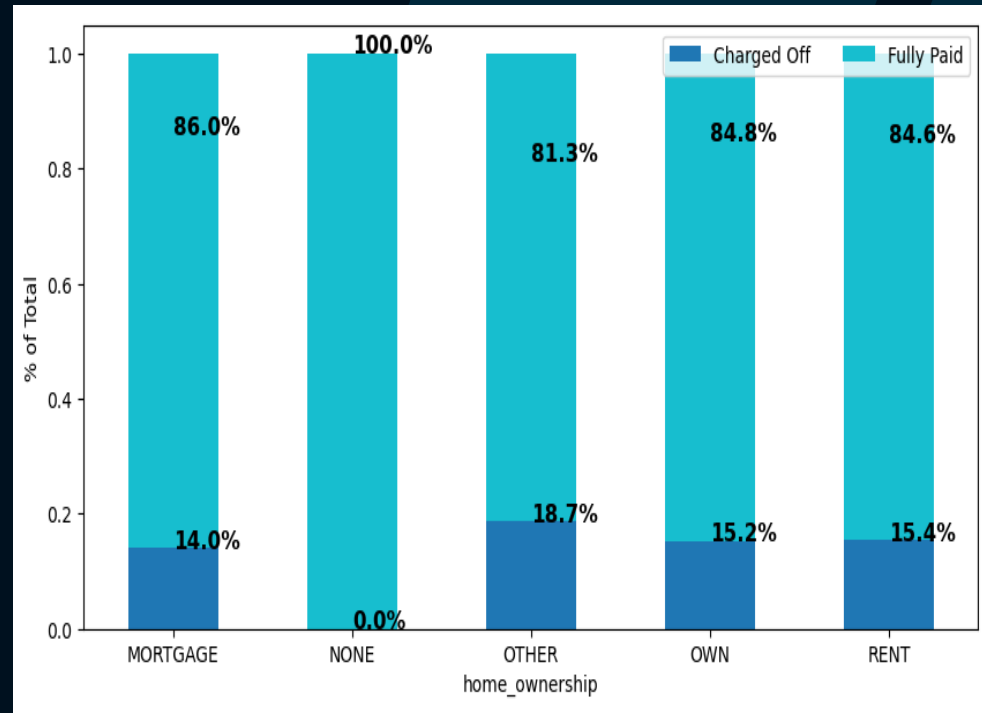
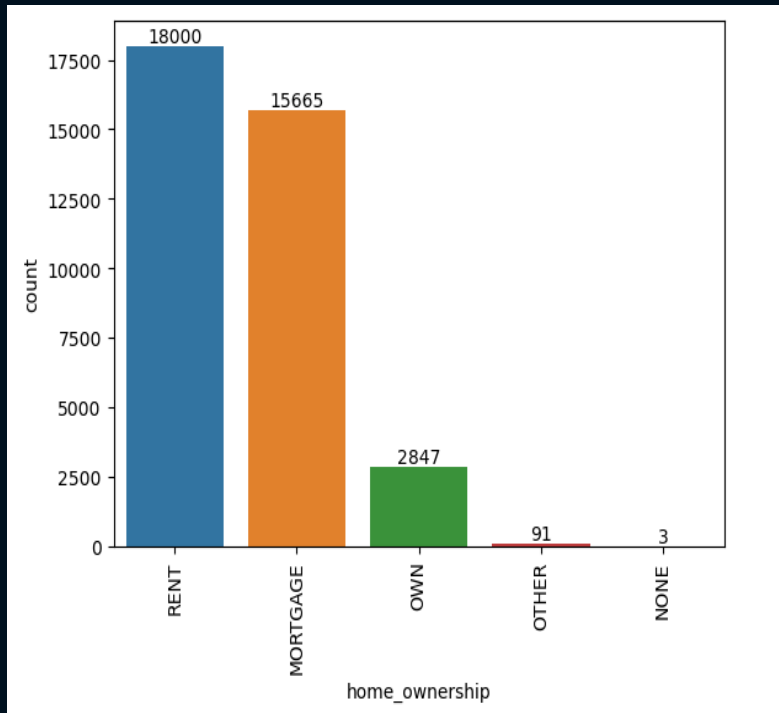
Analysis – Incidences of Delinquency



Observations

- Most of the loans are provided to borrowers with no history of delinquency
- Loans with more than 1 incidences of delinquency has higher rate of default

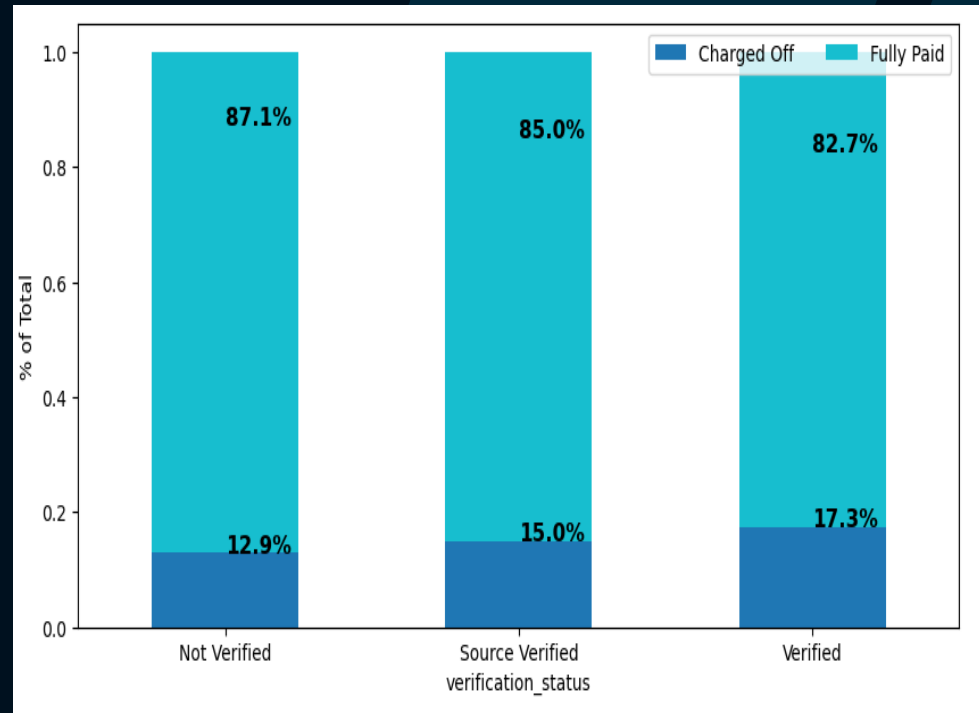
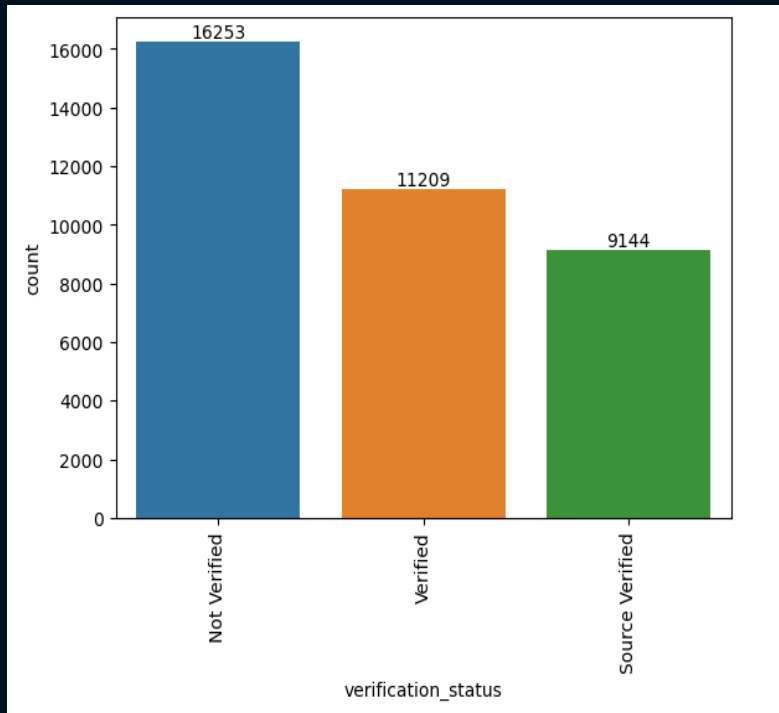
Analysis – Home Ownership



Observations

- Most of the loans are by people on rent or mortgage
- There is no significant impact of home ownership on default

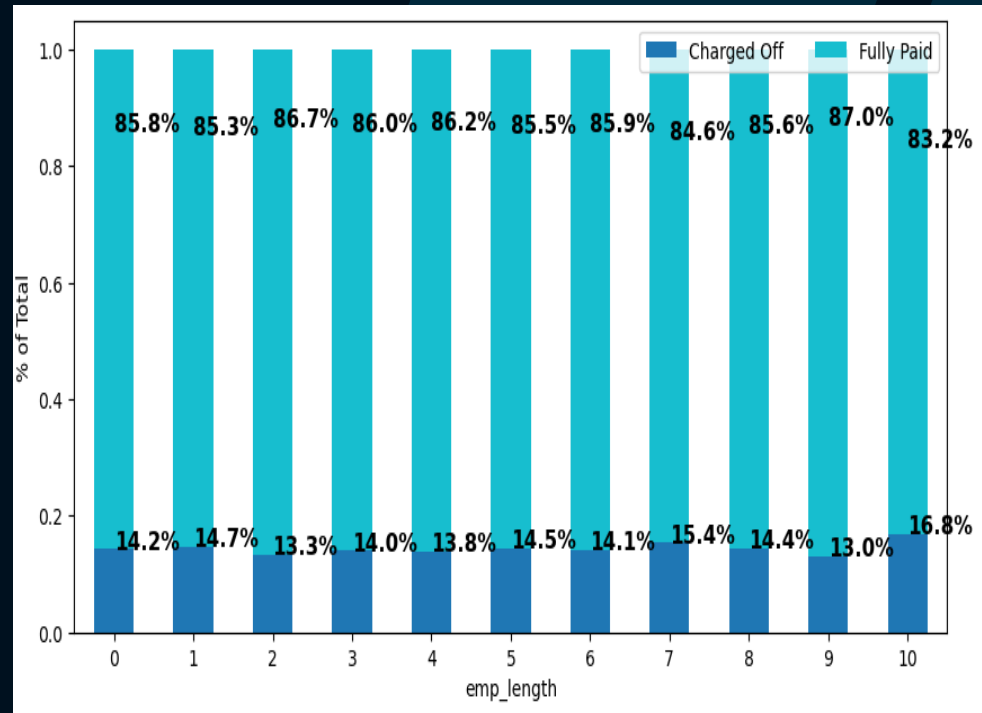
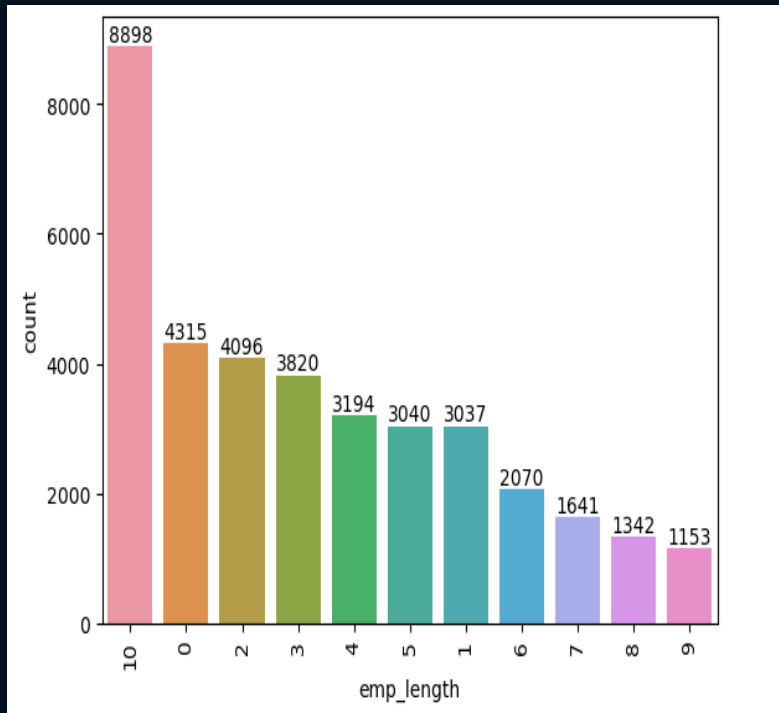
Analysis – Verification



Observations

- Almost 50% of the loans are not verified
- The chances of default for verified loans are more than not verified loans

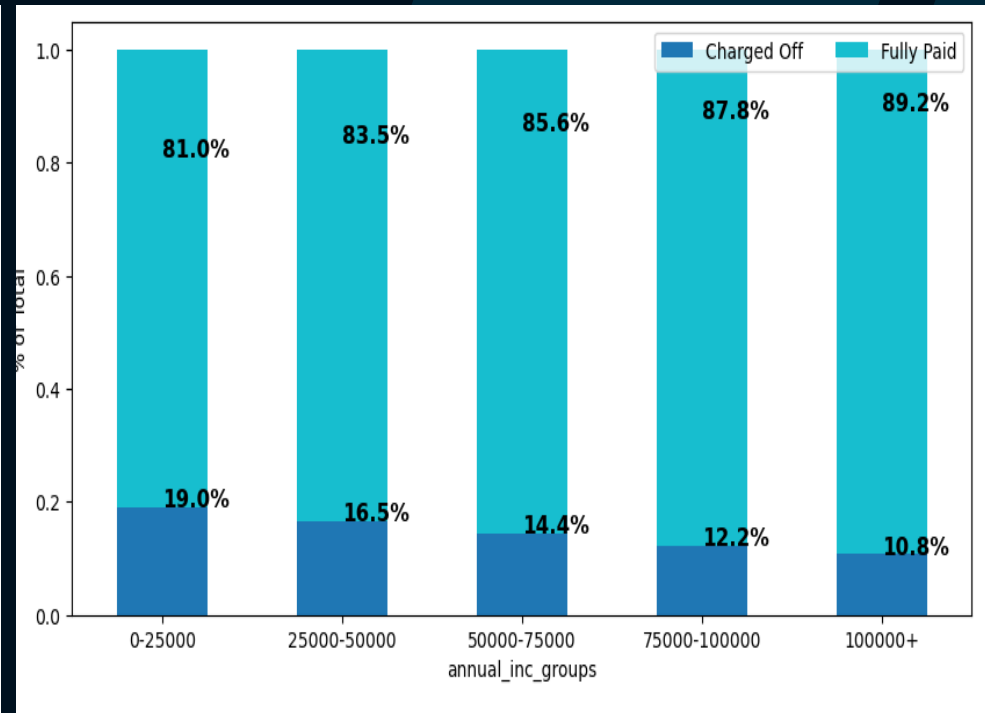
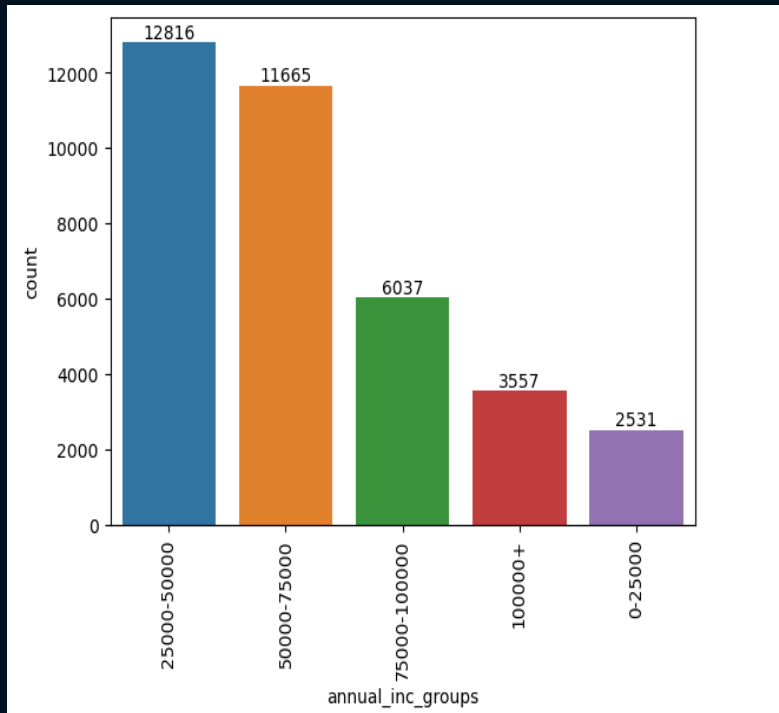
Analysis – Employee Length



Observations

- Most of the borrowers are of experience 10 years and above

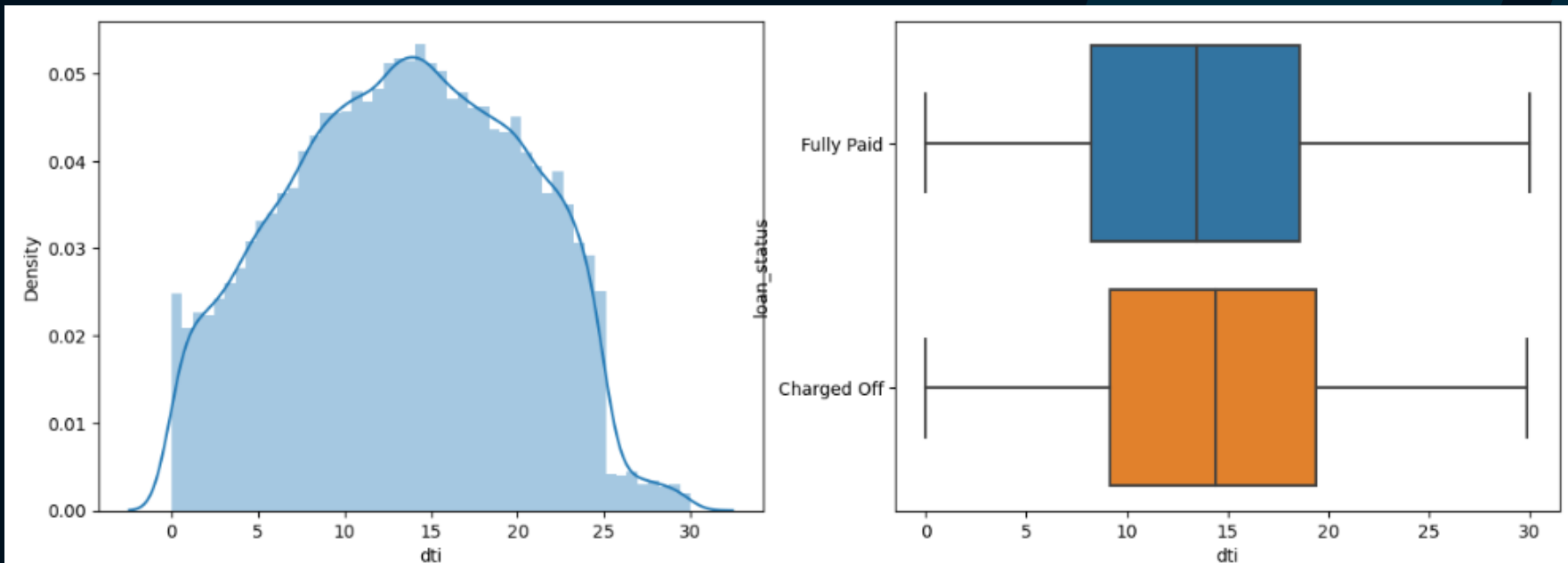
Analysis – Annual Income



Observations

- Borrowers with income less than 25K are more likely to default
- Loan default decreases with higher annual income

Analysis – DTI

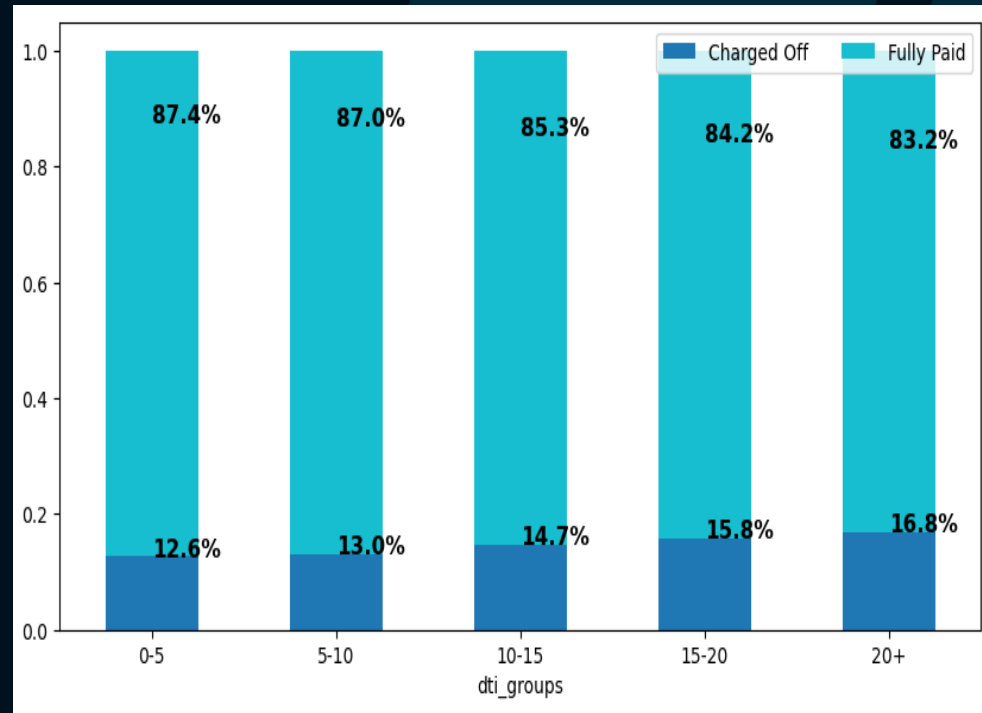
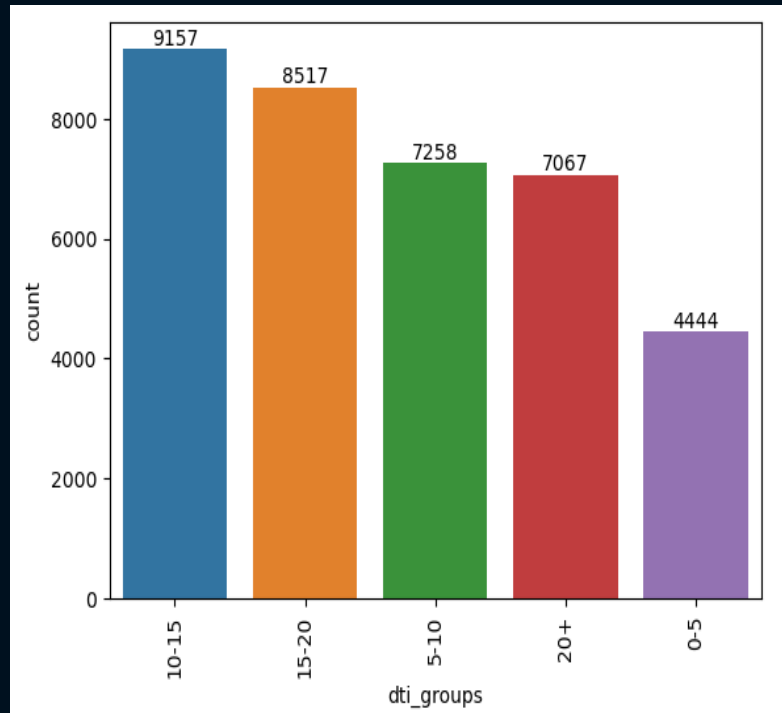


	count	mean	std	min	25%	50%	75%	max
loan_status								
Charged Off	5402.0	14.110500	6.586738	0.0	9.18	14.42	19.41	29.85
Fully Paid	31204.0	13.335826	6.666556	0.0	8.21	13.43	18.60	29.99

Observations

- Looks like higher DTI increases the likelihood of default
- Let's analyse more by creating bins

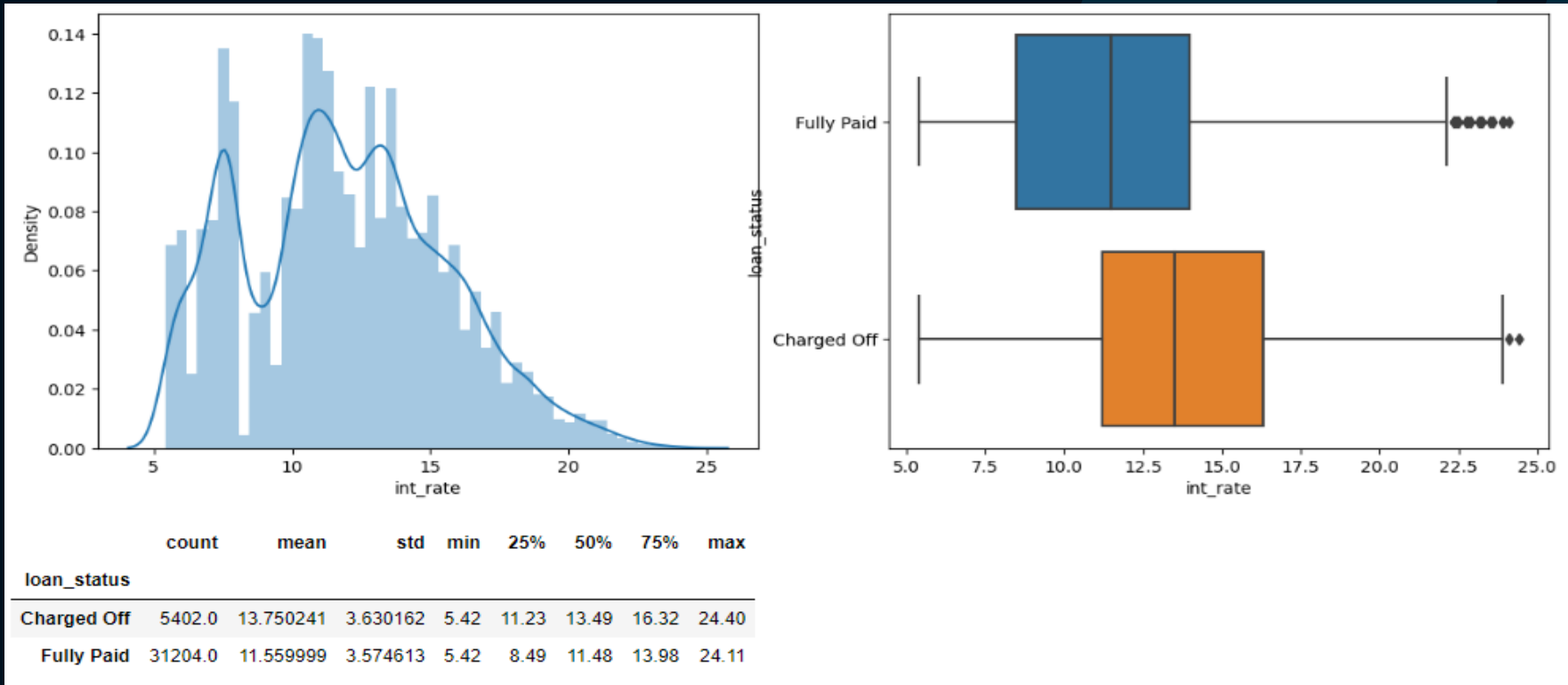
Analysis – DTI Continued



Observations

- Higher the DTI or debt to income ratio, higher the chances of loan being Charged Off

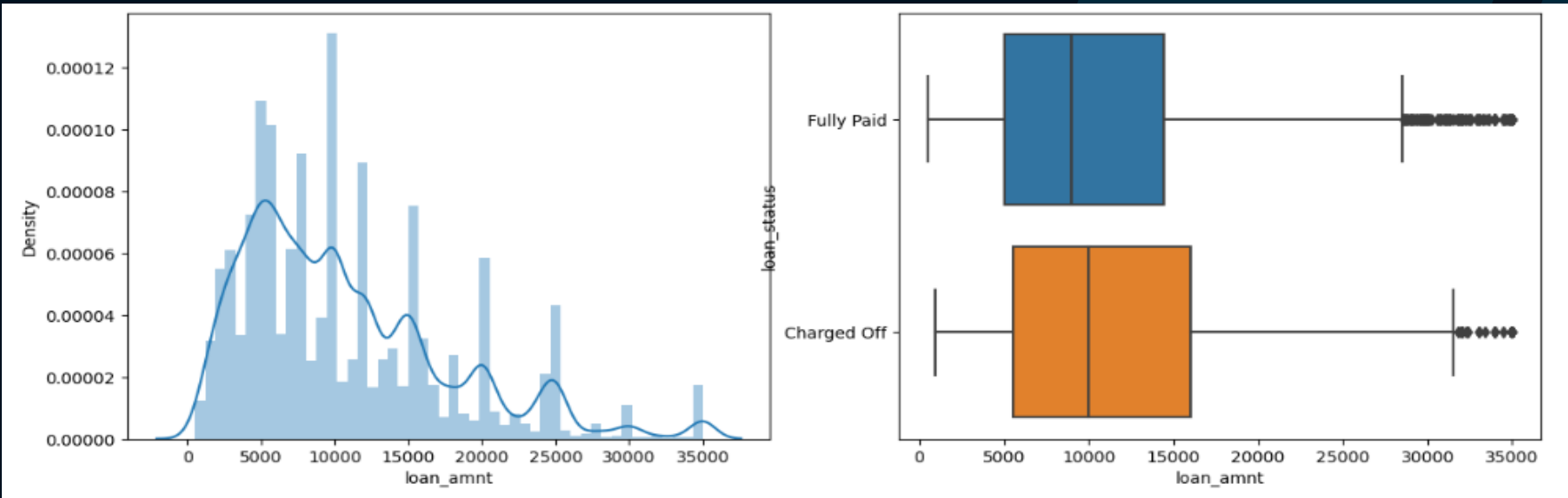
Analysis – Interest Rate



Observations

- Interest rates for defaulted loans are higher than fully paid ones. One inference can be that they are more likely to be the riskier loans

Analysis –Loan Amount

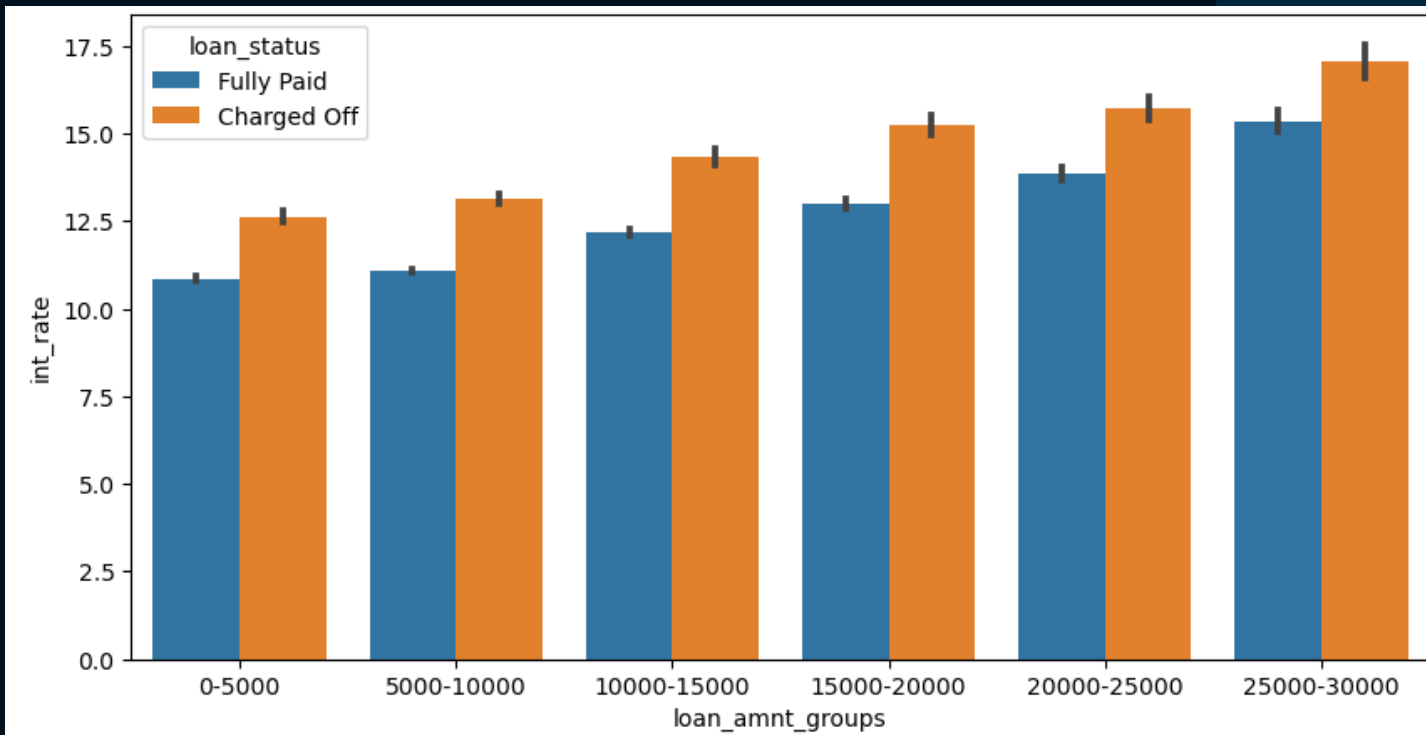


	count	mean	std	min	25%	50%	75%	max
loan_status								
Charged Off	5402.0	11831.307849	7885.005738	900.0	5500.0	10000.0	16000.0	35000.0
Fully Paid	31204.0	10492.486540	6874.206549	500.0	5000.0	9000.0	14400.0	35000.0

Observations

- The loan amount is right-skewed, where the median is less than the mean
- Defaulted loans are slightly on the higher side of the average loan amount

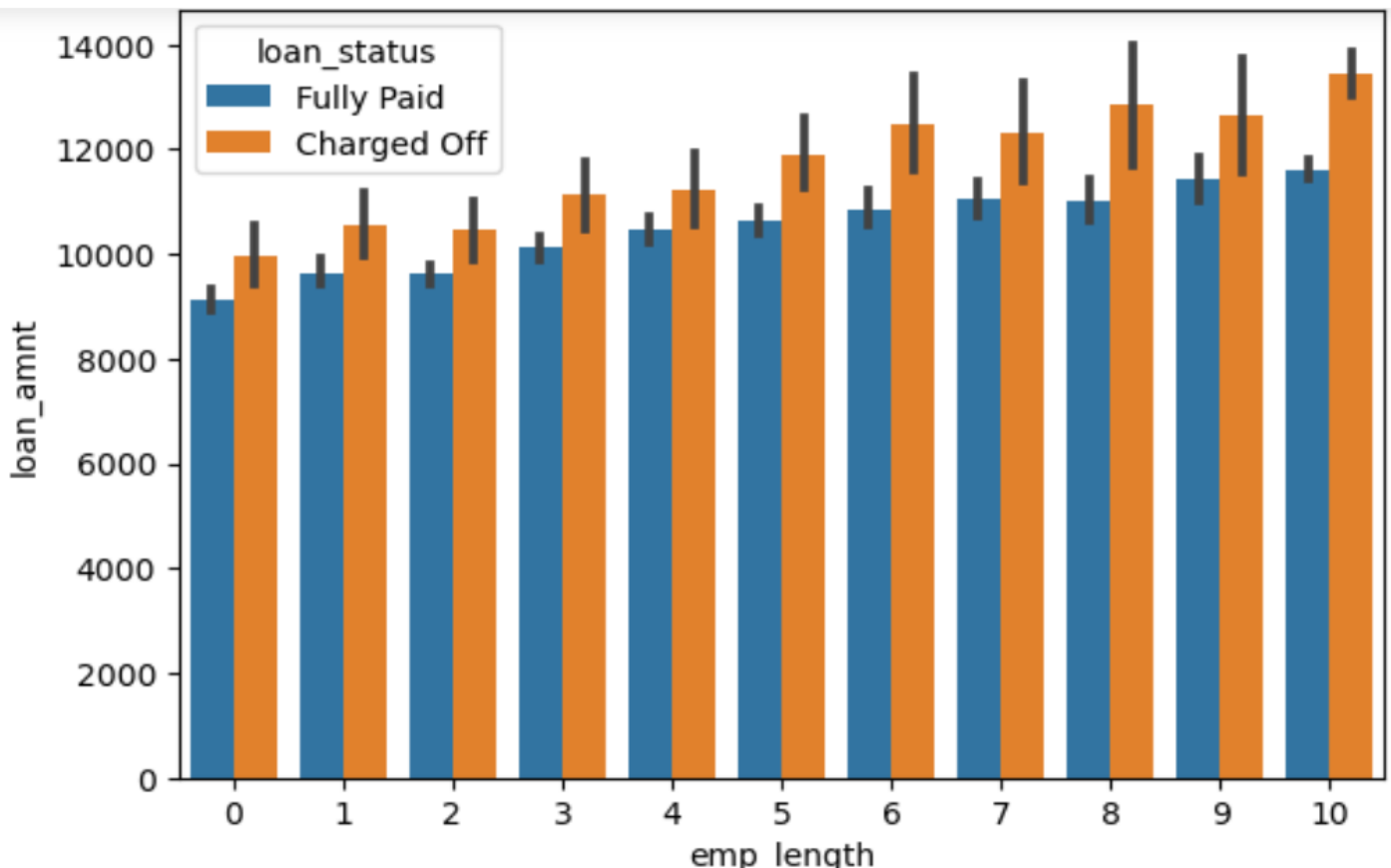
Analysis –Loan Amount vs Loan Status vs Interest Rate



Observations

- Higher the loan amount, higher will be the interest rate and chances to default
- There is a more probability of defaulting when applicants who have taken a loan in the range 30k - 35k and are charged interest rate of 15-17.5 %

Analysis –Loan Amount vs Employment Tenure



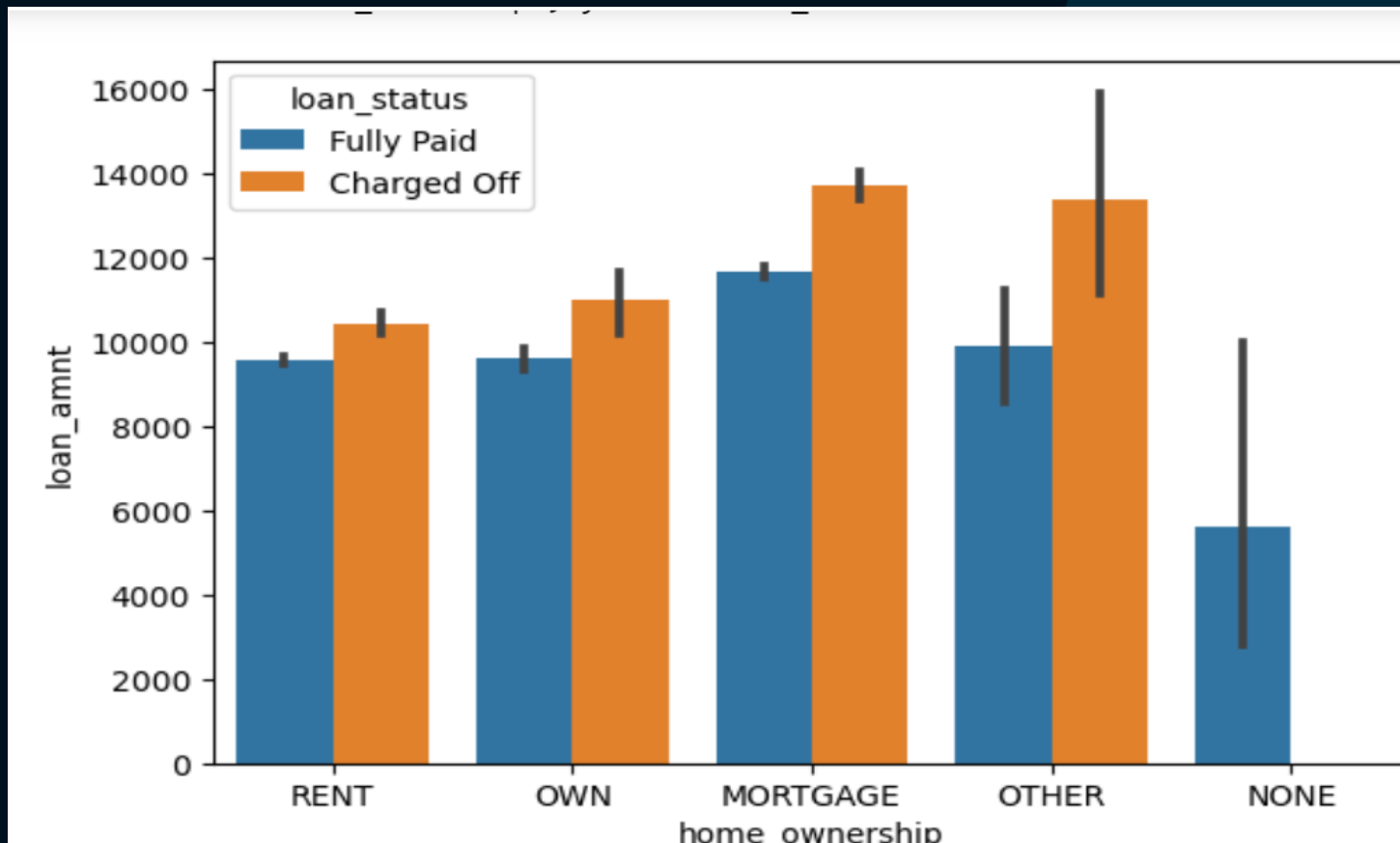
Observations

- Employees with tenure as 10 years usually takes loans of higher amount, and the risk is also low as compared to other tenure

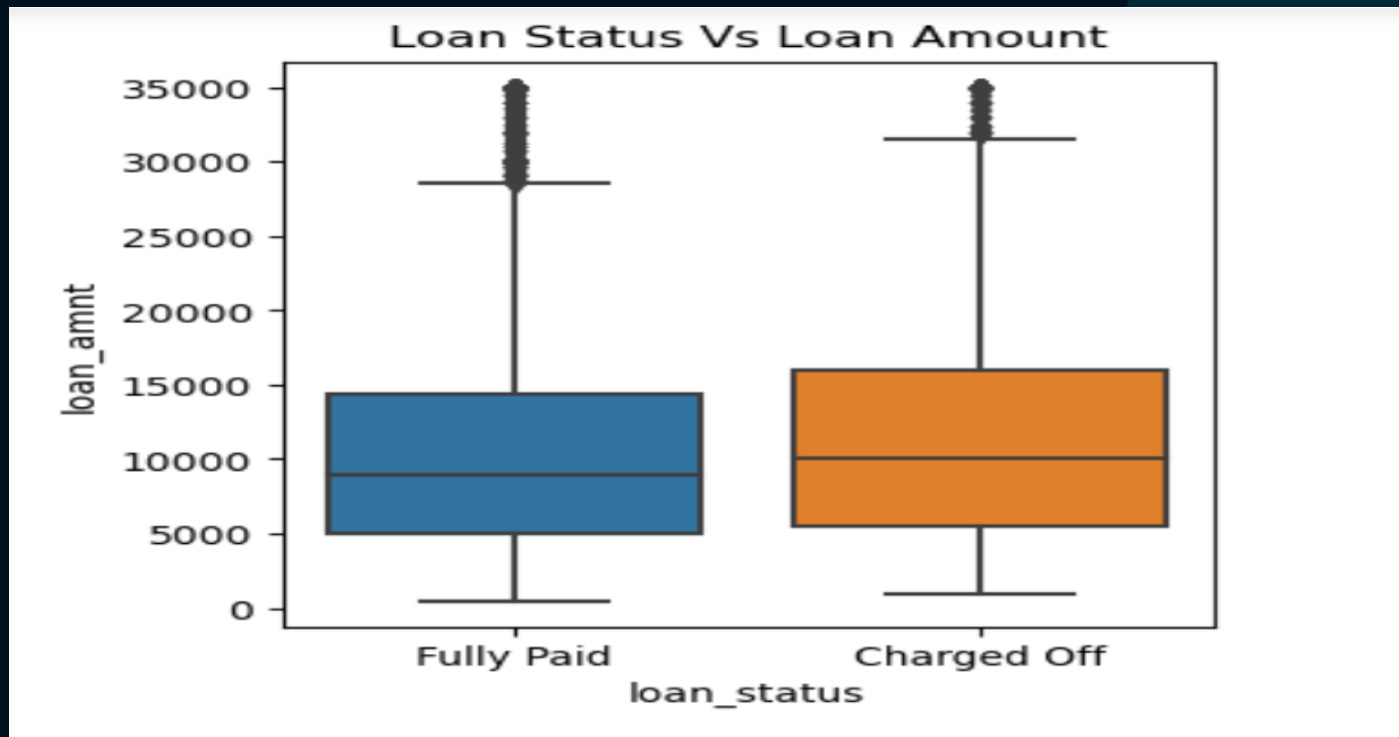
Analysis –Loan Amount vs Home Ownership

Observations

- Loan for Individuals living on rent usually have a higher chance of getting defaulted as compared to others. Applicants whose home ownership is 'MORTGAGE' are having more probability for getting loan charged off



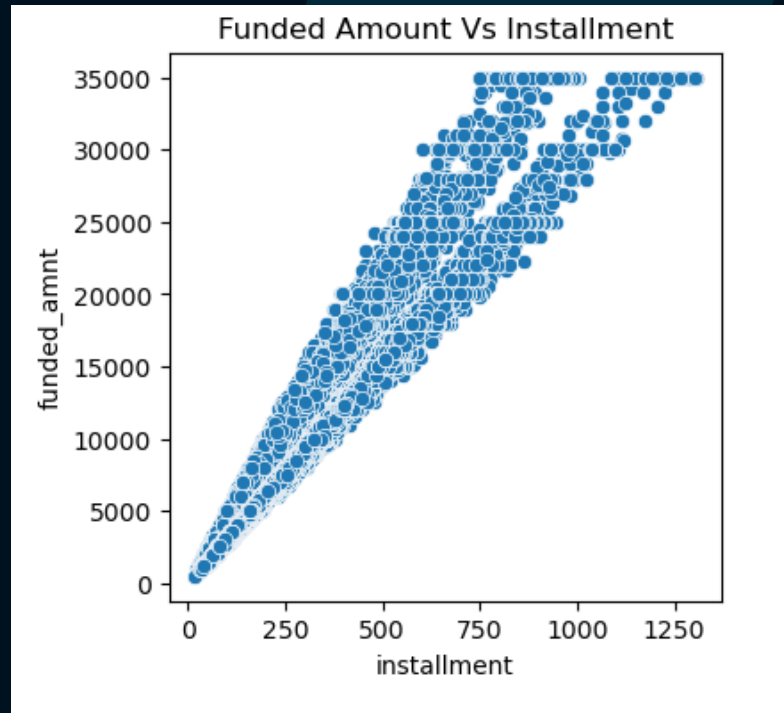
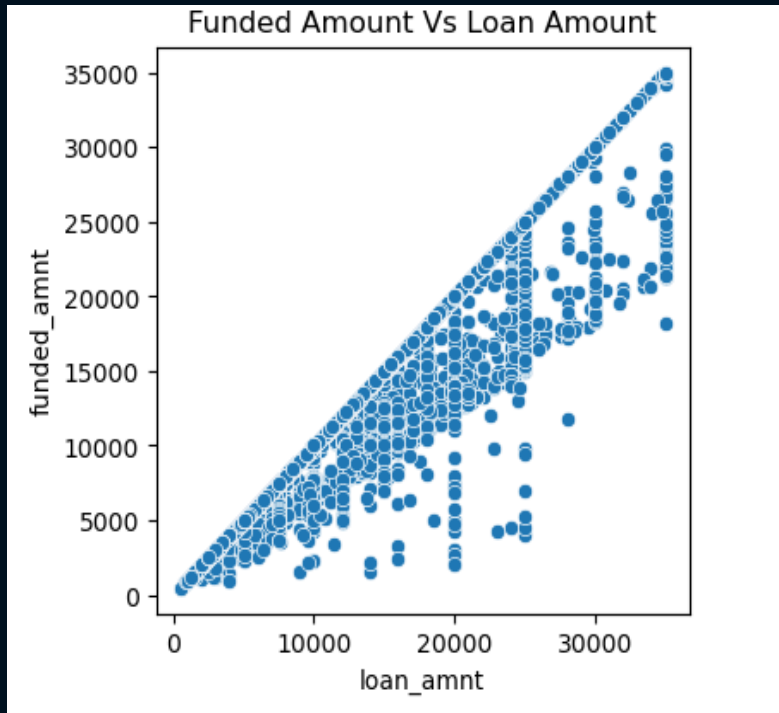
Analysis –Loan Status vs Loan Amount



Observations

- Chances of getting loan amount charged off is more than fully paid.

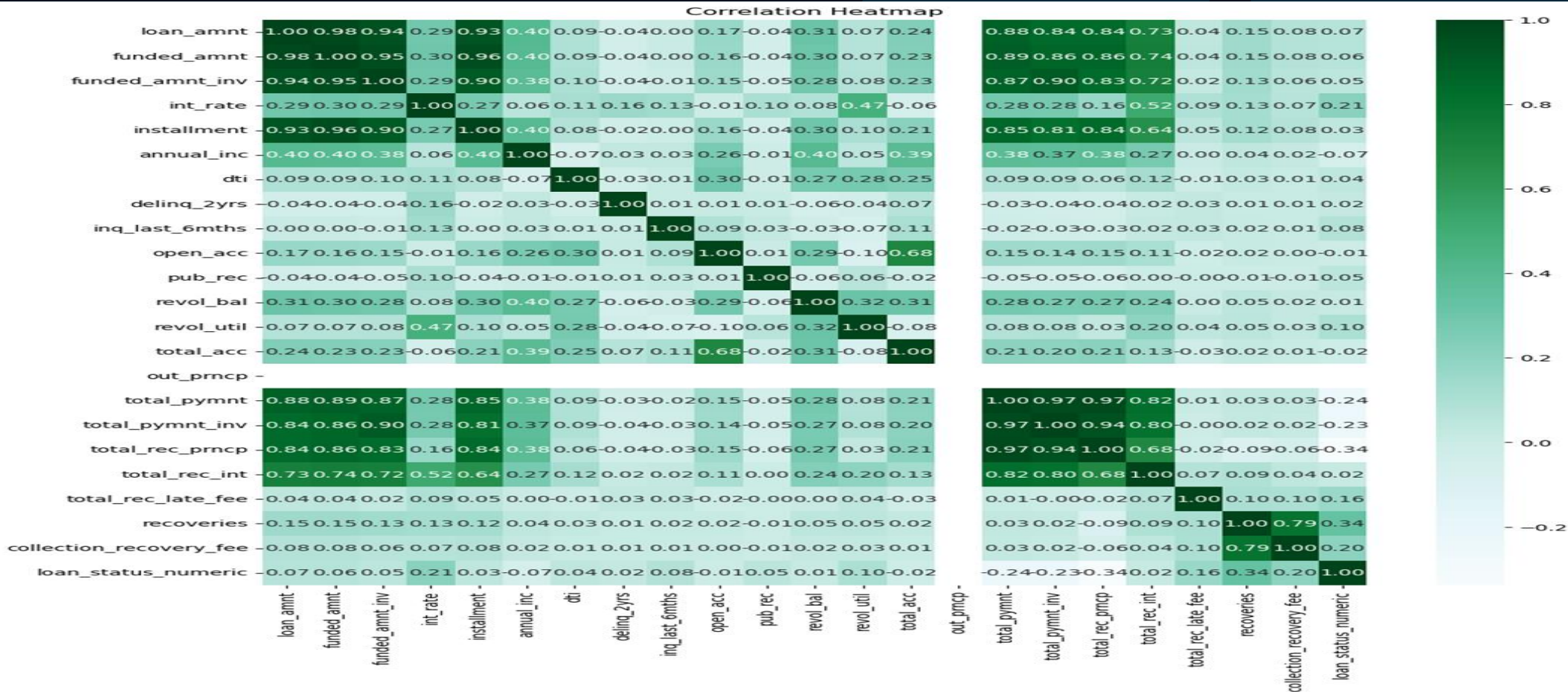
Analysis – Funded Amount vs Loan Amount vs Installment



Observations

- The funded amount, loan amount and installment are highly correlated
- Funded amount by the company is always less than or equal to loan amount applied by borrower

Correlation Analysis



Correlation Analysis

Observations

- Higher int_rate (0.21) can lead to high defaulters
- Higher recoveries (0.20) can lead to high defaulters
- Higher collection_recovery_fee (0.20) can lead to high defaulters
- Very high correlation between loan amount applied by the borrower and funded amount committed by the investor
- Positive correlation can also be found on funded_amnt, funded_amnt_inv, installment, dti, delinq_2yrs, inq_last_6mnths, pub_rec, revol_bal, revol_util, total_rec_int, total_rec_late_fee, pub_rec_bankruptcies
- High correlation between loan amount, monthly installment and principal received are highly correlated.
- Number of open credit lines is also highly correlated with total number of credit lines in the borrower's credit file.
- Revol_util, int_rate - Moderately high correlation
- Higher the amount of credit the borrower is using, higher is the risk of loan repayment. That's why higher is the interest rate, and vice-versa.

Recommendations

- Lending Club (LC) should keep on approving more loans of A or B grade than other grades
- LC should be careful before lending to small businesses as they have high default rate
- LC should avoid approving borrowers with prior public record or incidence of delinquency
- LC should charge higher rates to borrower with 20+ debt to income ratio as they are more likely to default
- LC should prefer 3 years loan over 5 years, as they are less likely to default





THANK YOU!

Subhomoy Ganguli
Swati Gupta Jain

Email

subhomoyganguli1001@gmail.com

swatigupta0106@gmail.com