



# Lending Club Case Study

BY

SALINI JOY

PRAVEEN KUMAR R

# Problem Statement

## Business Understanding

- ▶ We work in a consumer finance company where we lend loans for various types of urban customers
- ▶ Company has to take decision on loan approval based on applicants profile.
- ▶ We have data set which has details about past loan applicants and whether they 'defaulted' or not.
- ▶ The aim is to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

# Business Objectives

- ▶ Lending loans to 'risky' applicants is the largest source of financial loss .
- ▶ Borrowers who default cause the largest amount of loss to the lenders. In this case, the customers labelled as 'charged-off' are the 'defaulters'.
- ▶ The company can utilize this knowledge for its portfolio and risk assessment.

# Dataset Understanding

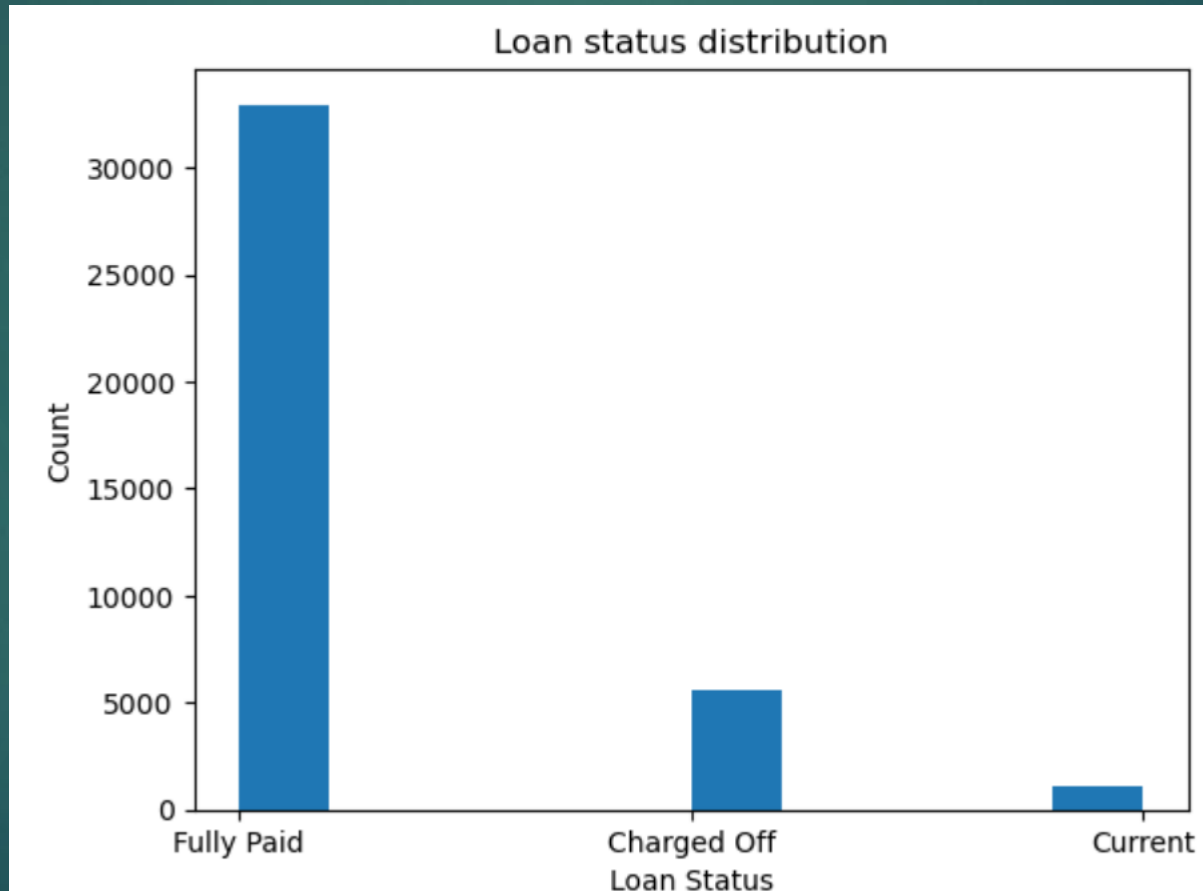
When a person applies for a loan, there are two types of decisions that could be taken by the company:

1. Loan accepted: If the company approves the loan, there are 3 possible scenarios described below:
  1. Fully paid: Applicant has fully paid the loan (the principal and the interest rate)
  2. Current: Applicant is in the process of paying the instalments, i.e. the tenure of the loan is not yet completed. These candidates are not labelled as 'defaulted'.
  3. Charged-off: Applicant has not paid the instalments in due time for a long period of time, i.e. he/she has defaulted on the loan
2. Loan rejected: The company had rejected the loan (because the candidate does not meet their requirements etc.). Since the loan was rejected, there is no transactional history of those applicants with the company and so this data is not available with the company (and thus in this dataset)

# Analysis Approach

- ▶ Getting the complete dataset information
- ▶ Drop columns having very high percentage null values
- ▶ Drop columns with single values
- ▶ Check for duplicate rows and drop it
- ▶ Filtering dataset to get Charged Off customers (Defaulters)
- ▶ Outlier treatment
- ▶ Univariate Analysis
- ▶ Bivariate Analysis
- ▶ Correlation Analysis
- ▶ Suggestions

# Loan status distribution



# Outlier Treatment

There is outliers in the annual income column. After outlier treatment the extreme variables have been removed .



# Univariate Analysis

Univariate analysis deals with analyzing variables one at a time. It is important to separately understand each variable before moving on to analyzing multiple variables together.

## Ordered categorical variables

- Grade (grade)
- Sub grade(sub\_grade)
- Term (36 / 60 months) (term)
- Employment length (emp\_length)
- Issue year (issue\_d\_y)
- Issue month (issue\_d\_m)
- Delinquency (delinq\_2yrs)

## Ordered categorical variables

- Address State (addr\_state)
- Loan purpose (purpose)
- Home Ownership (home\_ownership)
- Verification status (verification\_status)

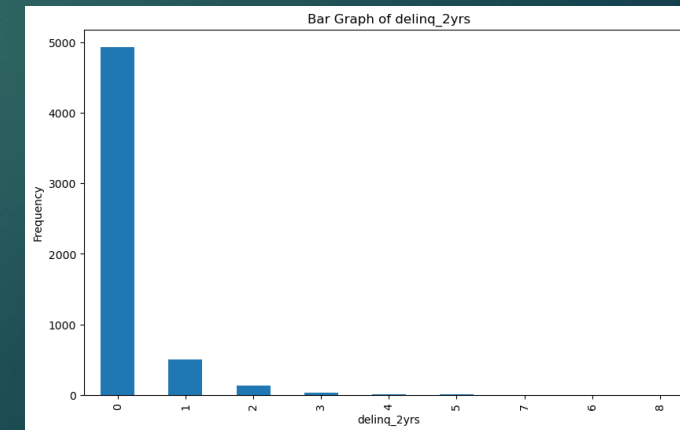
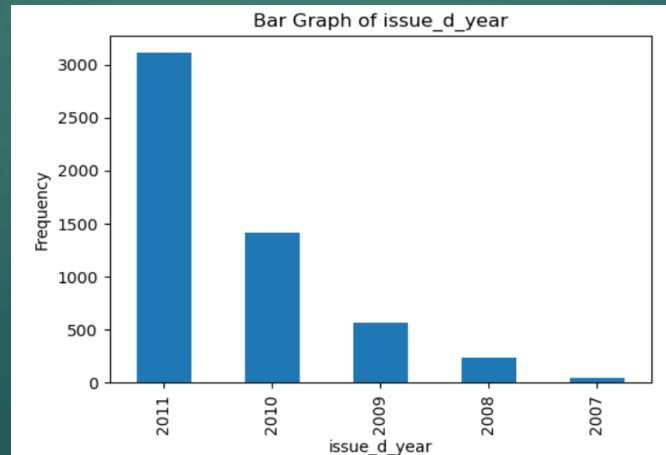
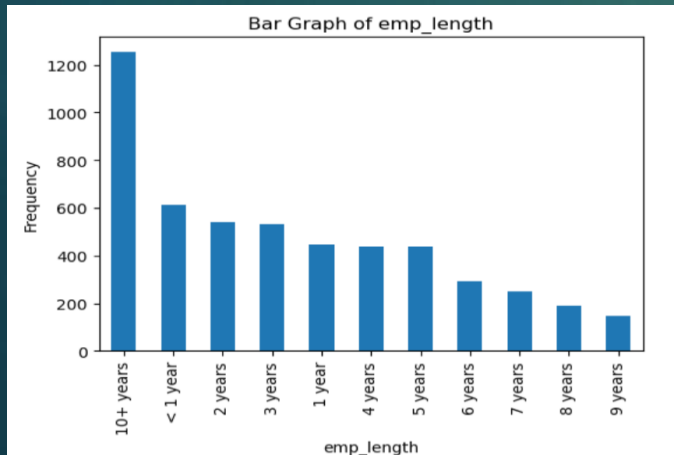
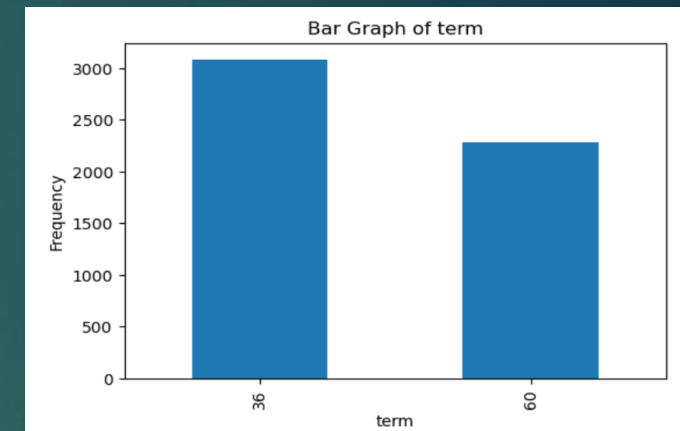
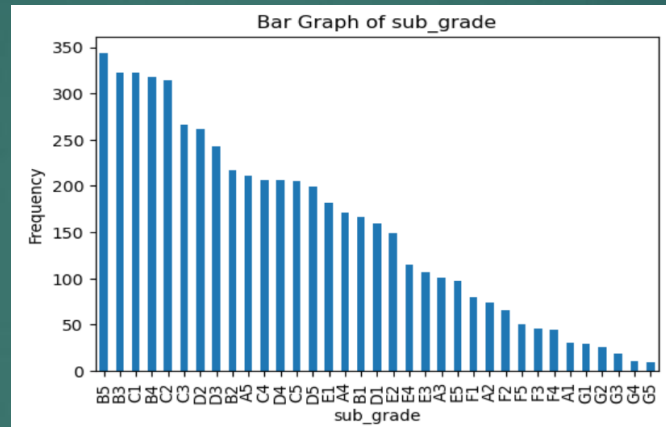
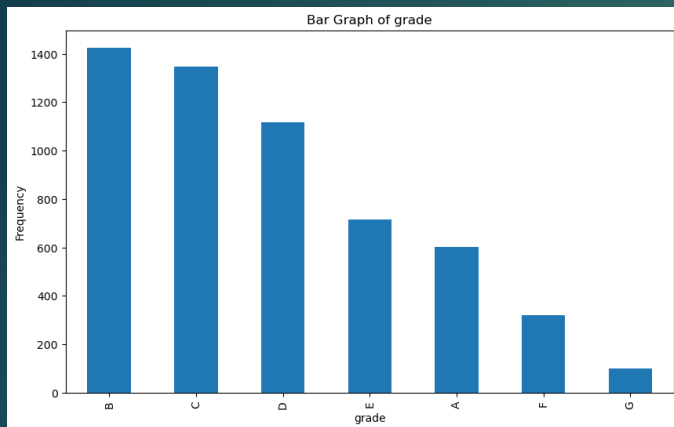
## Quantitative variables

- Annual income (annual\_inc\_bins)
- Loan amount (loan\_amnt\_bins)
- Interest rate (int\_rate\_bins)
- Funded amount (funded\_amnt\_bins)
- Debt-to-Income Ratio (dti\_bins)
- Instalment (installment\_bins)



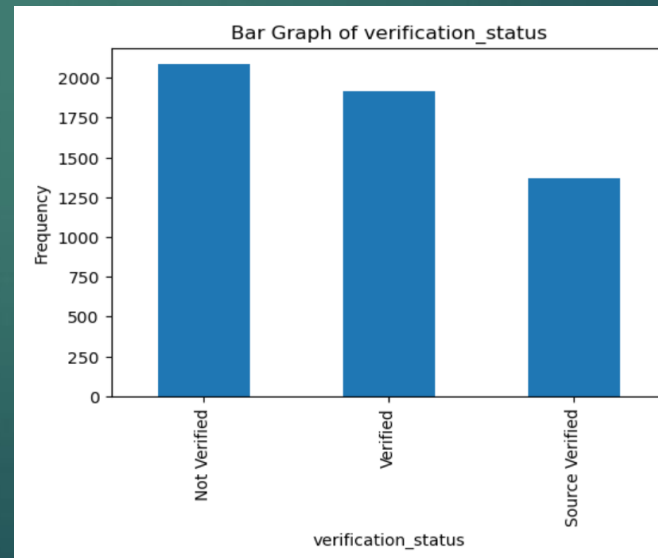
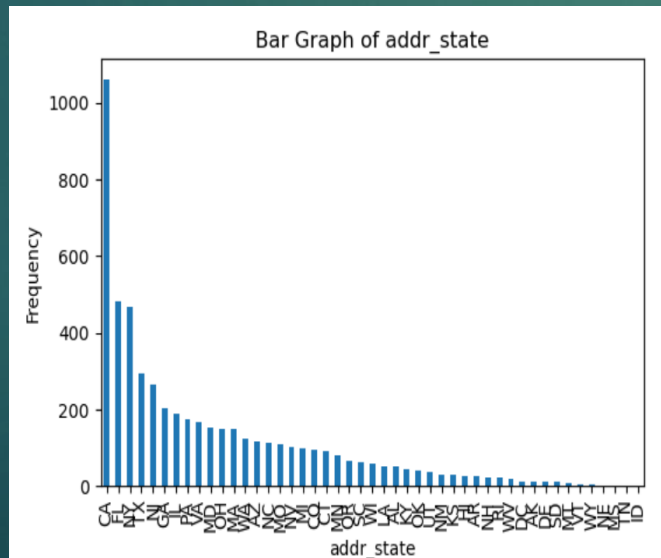
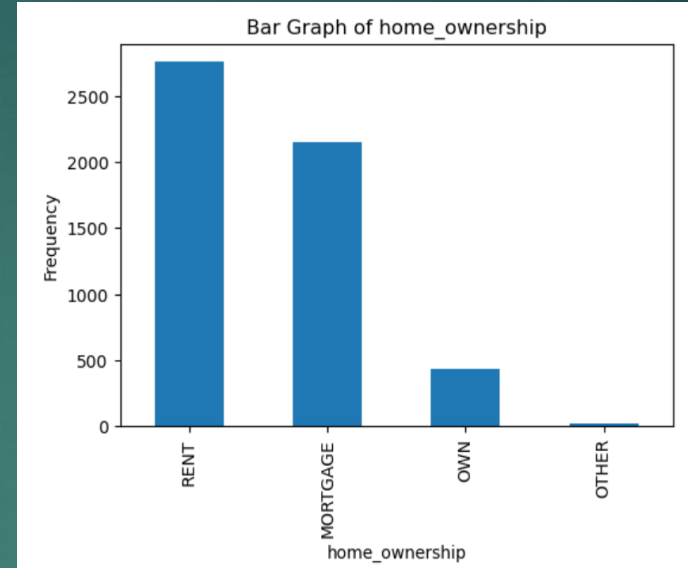
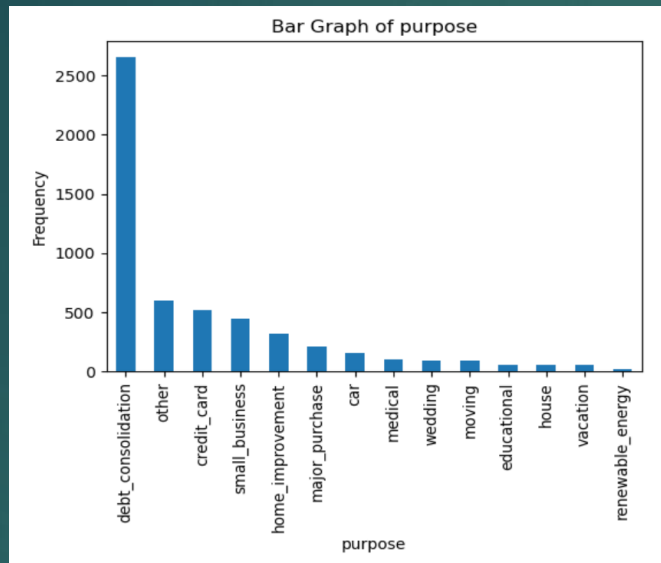
# Univariate Analysis

## Ordered categorical variables of defaulters



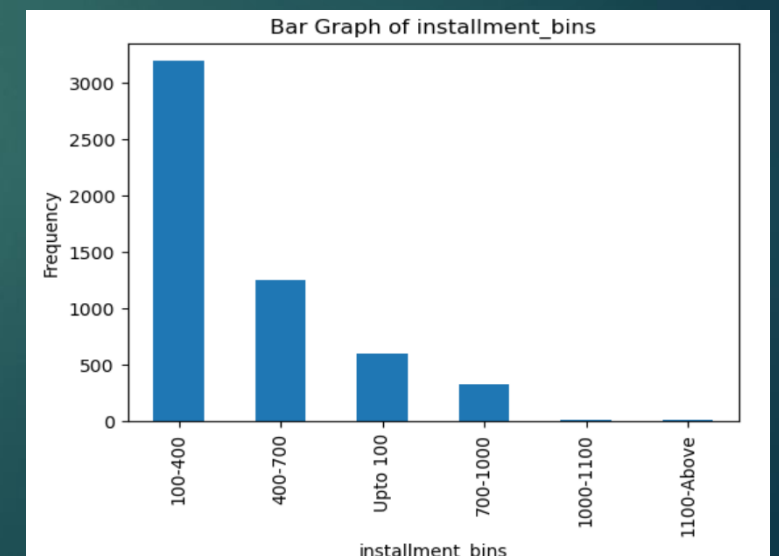
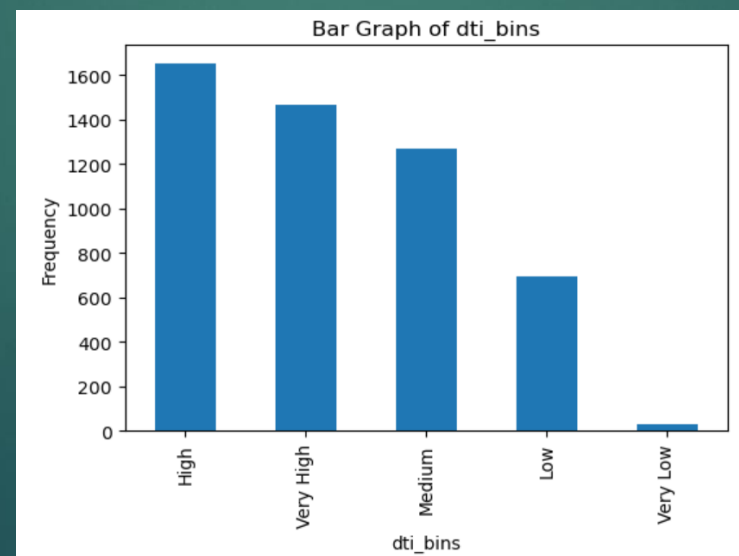
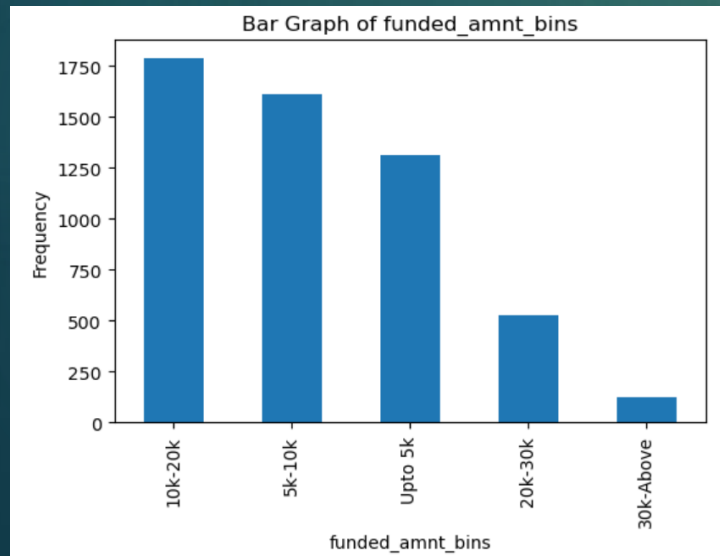
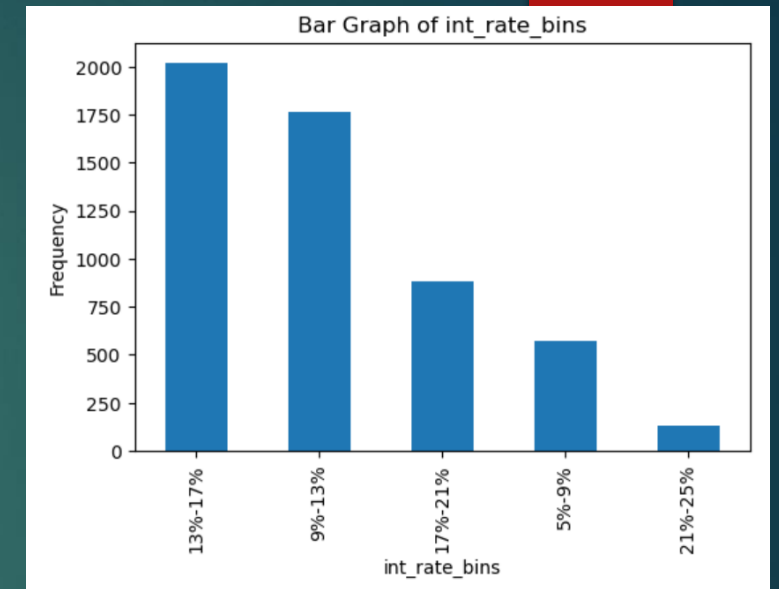
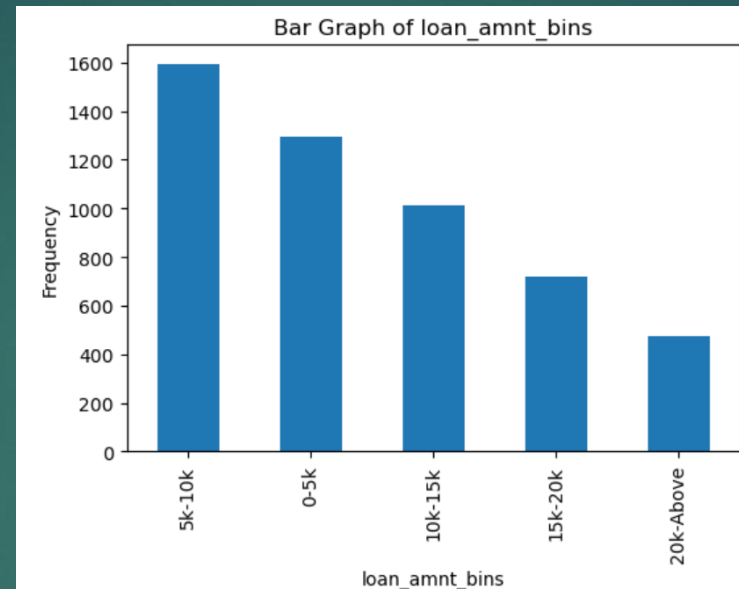
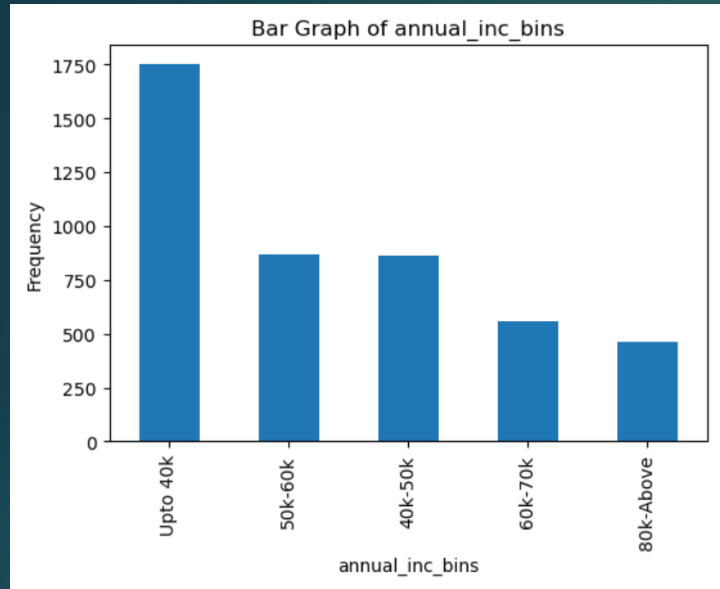
# Univariate Analysis

## Unordered categorical variables of defaulters

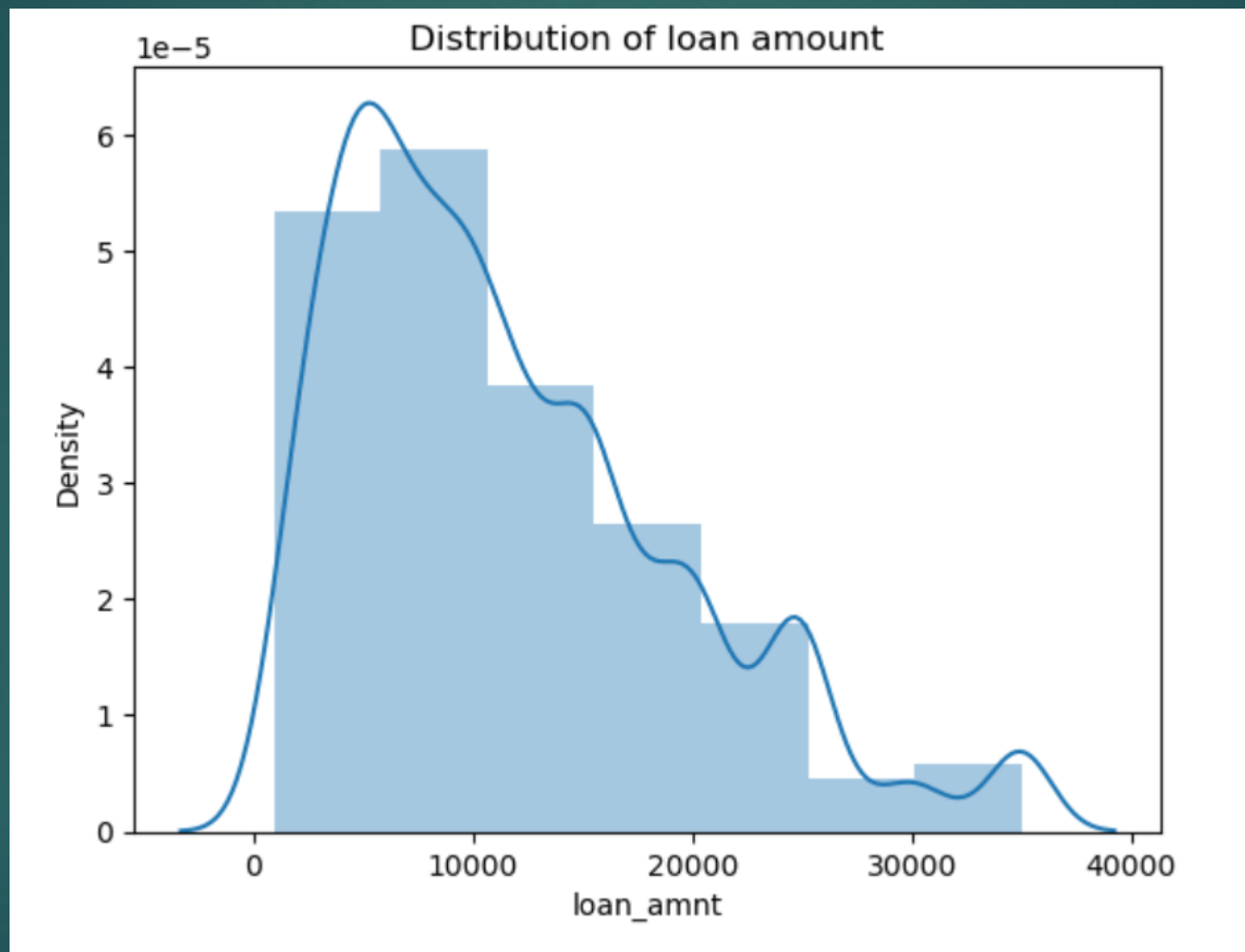


# Univariate Analysis

## Quantitative variables of defaulters



# Distribution of loan amount



The maximum loan amount range is between 5000 and 10000

# Observations from Univariate Analysis

## From the ordered categorical variable analysis:

- Grade B has the highest number of defaulters followed by Grade C.
- Among Grade B customers, sub grade B5 is the highest number of defaulters. Not all B grade applicants are likely to be a defaulter.
- Loan applicants with term 36 is more than term 60. The defaulters chose less term. So the the amount per month was more are that could be the reason for making them defaulters
- The customers who have been working for more than 10 years are the highest number of defaulters. This suggests that even though they have a permanent income they are making dues in payments. So the employment period is not a great point to consider the applicant is likely to pay back the loan amount. The second highest is the customers having employment less than 1 year.
- Highest number of loan was funded in the year 2011.
- Defaulted loans were funded in December. Last quarter shows the highest number which got defaults. We can conclude loan taken during the last quarter is most likely to get defaulted.(The last quarter is the Holiday season)

## From the unordered categorical variable analysis:

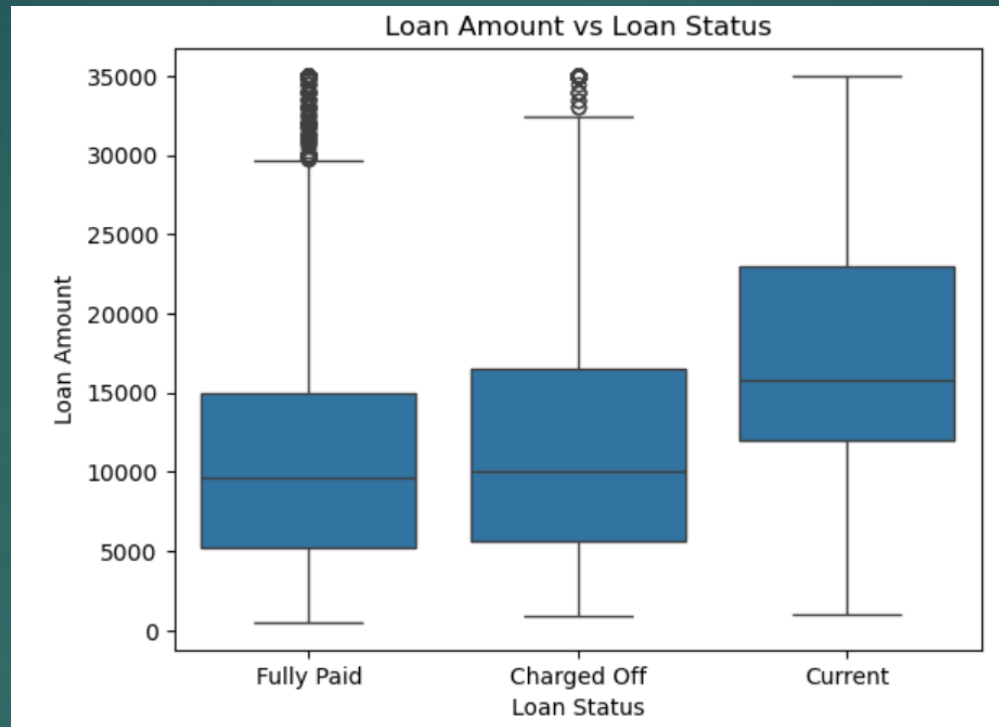
- Debt consolidation is the prime purpose for the defaulters to take the loan. They were already having financial issues and took new loan. This needs to be considered very cautiously while approving the loan.
  - For most of the defaulters, their home ownership is Rent. This suggest the defaulters are already having financial issues.
  - Majority of the defaulters are from California. The lending company needs to verify the details of this loan applicants thoroughly.
  - Lening company has approved more loans without proper verification.
- Verification status indicates if income was verified by LC, or not verified, or if the income source was verified. We can also see that verified status defaulters are also not less. Verification process needs to be more efficient and this will help in reducing the risk of funding to defaulters.

## From the numerical variable analysis:

- The maximum defaulters have annual salary below 40000. The loan was funded very less to the applicants having annual income less than 20000. This is appreciable. But in order to reduce the charged off situation the source verification will be more helpful.
- Majority of the defaulters received a loan amount in the range 5000 -10000. The second group of defaulters were funded a loan amount less than 5000. Even though the amount was less, they were not able to repay it on time.
- While checking the interest rate we can see that 13%-17% is the highest. The lending company should work on the interest rate that is being applied on the loan amount. Defaulters were not able to make this high interest rate amount.
- It is observed that many defaulters had received a funded amount in the range 10000-20000. Lending company should approve this after proper verification of financial status of the loan applicants.
- Many of the defaulters has high dti ratio. Lending company should consider this parameter for approving the loan.



# Comparison of Fully paid, Current and Charged off customers

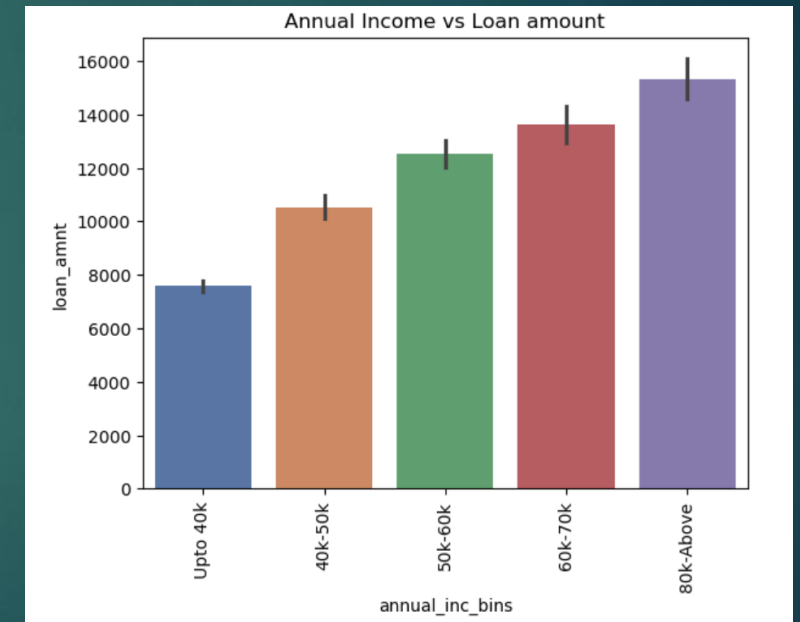
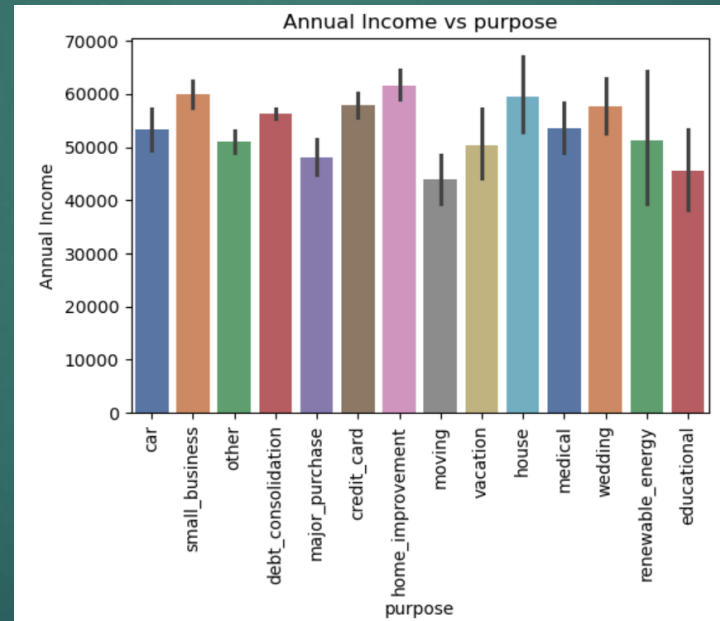
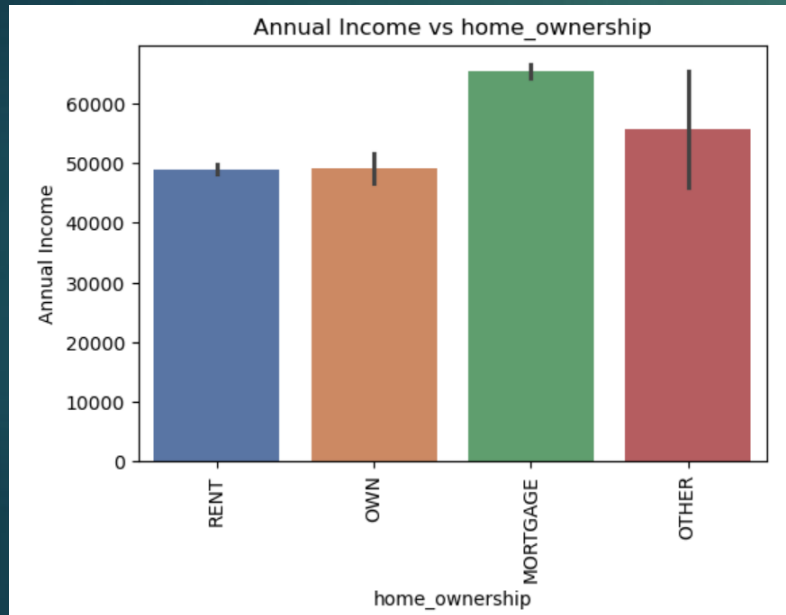


- From the box plot of fully paid and Charged off customers we can see similar patterns. Large loan amount is more likely to default. The most of the defaulters have the loan amount in the range 5000-15000. There are few outliers in the defaulters.
- Rent and Mortgage customers are the highest in case of both fully paid and charged off.
- Customers having 10+ years of experience is the highest for fully paid and charged off.

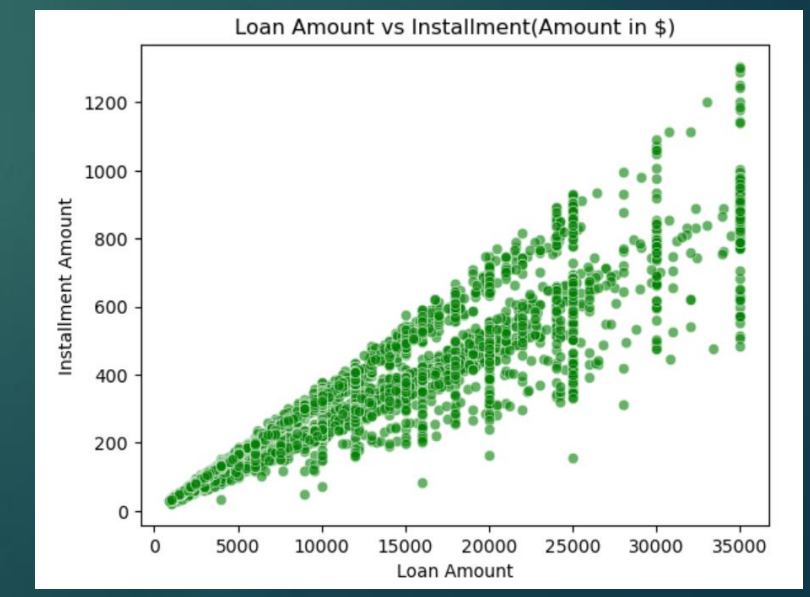
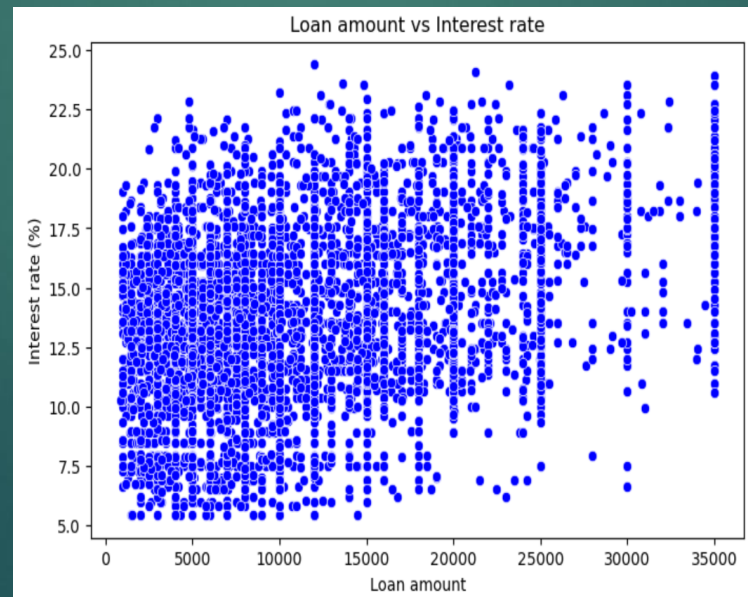
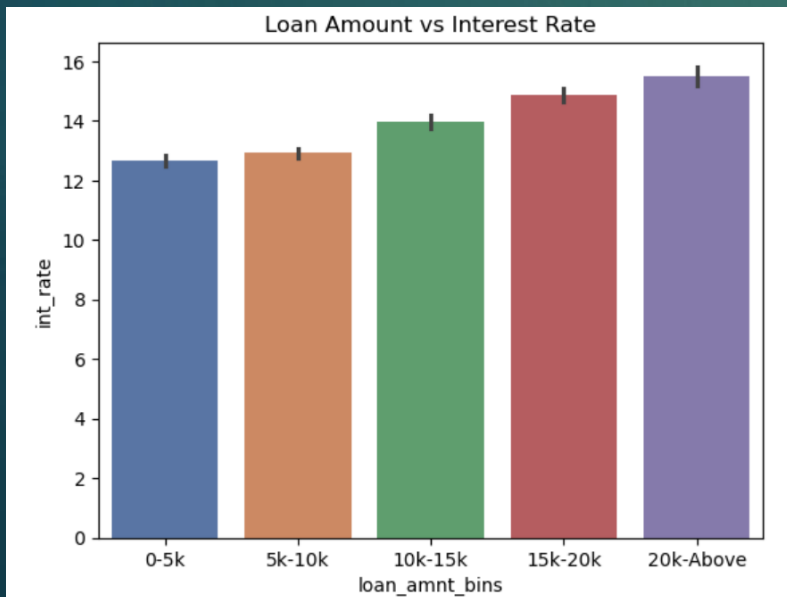
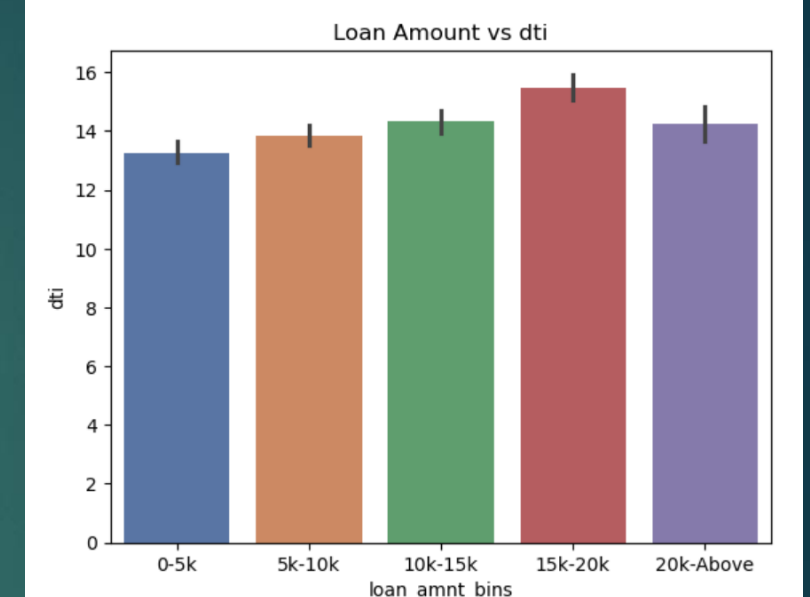
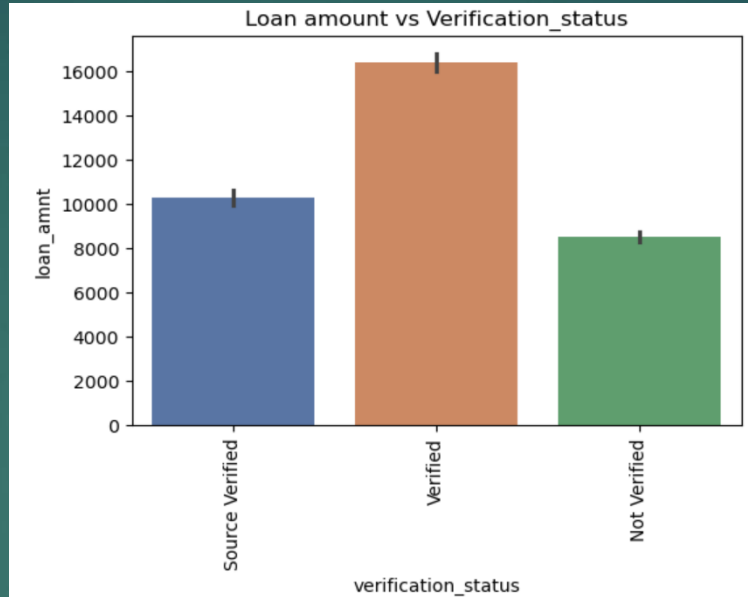
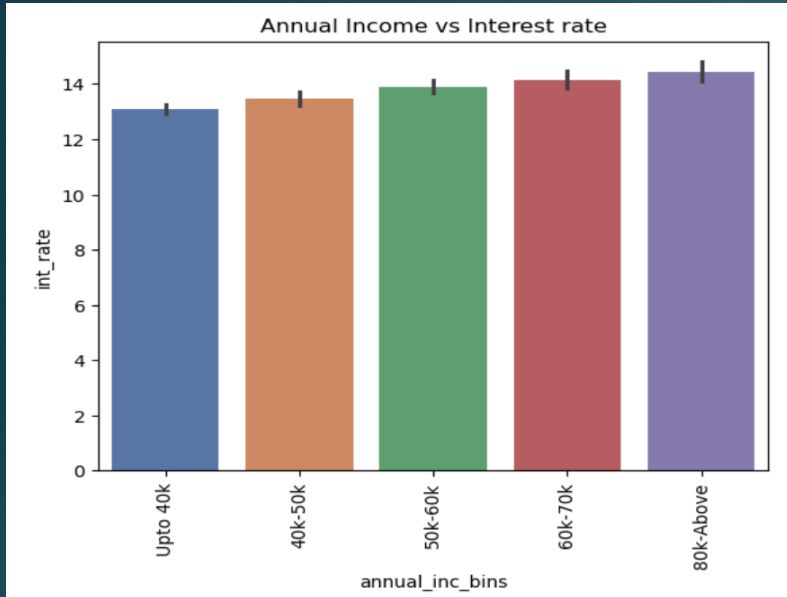


# Bivariate Analysis

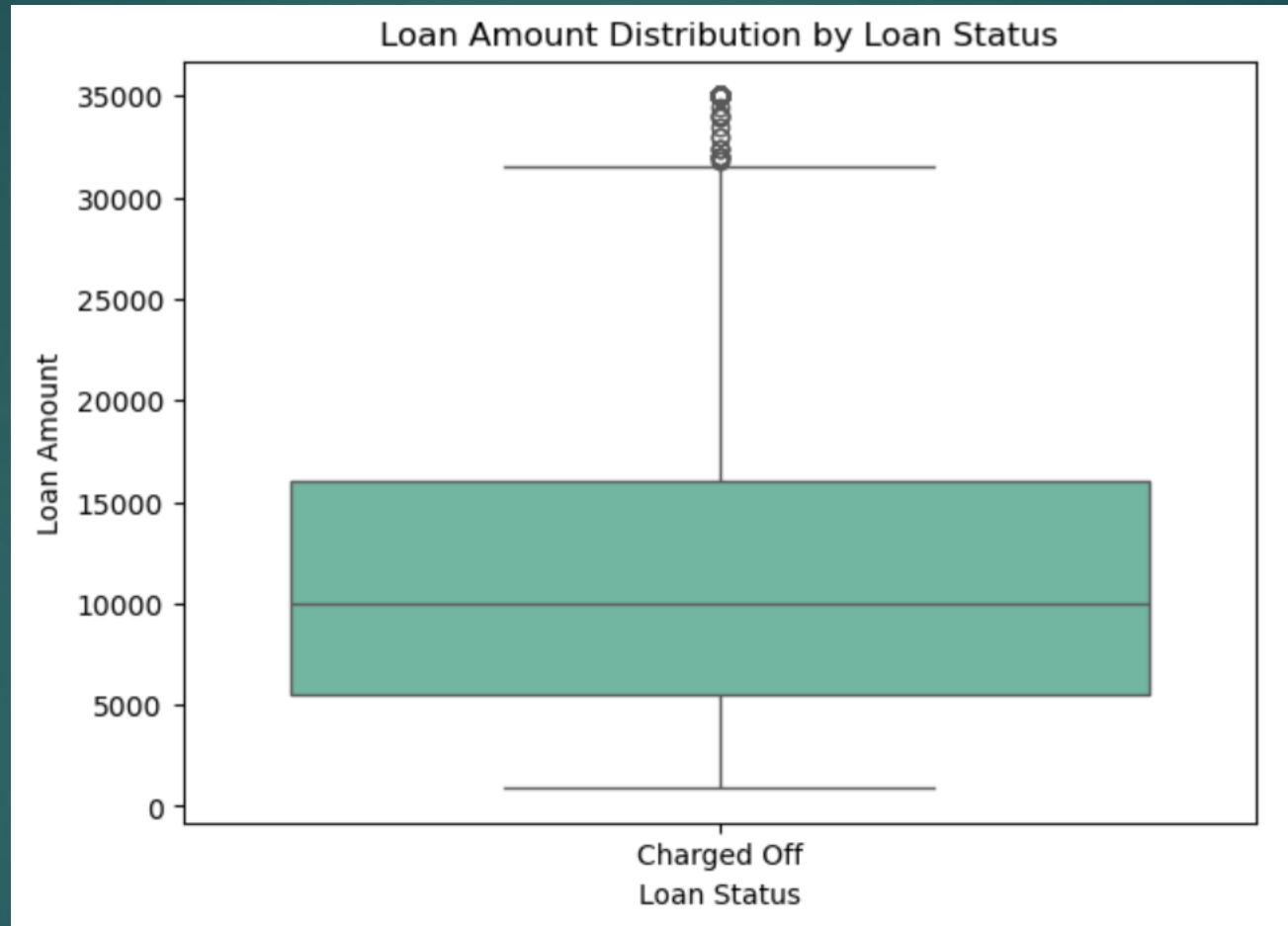
Bivariate analysis is a statistical method used to analyze the relationship between two variables. It helps determine whether and how two variables are related, and if so, the strength and nature of this relationship.



# Bivariate Analysis



# Loan Amount Distribution by Loan Status



# Observations from Bivariate Analysis

- It is observed that the defaulters who have higher annual salary took loan for home improvement followed by house, small business and wedding
- More loan amount was funded for small business
- Loan applicants having employment length more than 10 years were funded highest loan amount
- Loan was approved mostly for verified applicants. This is a good trend. This will reduce the risk of charged off customers.
- Maximum average loan amount was funded to the applicants who opted term 60.
- Most of the loan applicants who charged off have significantly high Debt-to-Income (dti) ratios.
- Majority of loan applicants who charged off received loans with interest rates in the range of 13% to 17%.
- Larger loan amounts are associated with higher interest rates for charged-off loans. This could be due to the greater risk associated with lending larger amounts.
- Loan applicants belongs to Grades F and G are more likely to be defaulters
- Loan amount funded increased from 2007 to 2011
- Lending company funded more loan amount to Verified loan applicants. This indicates that company is very cautious in approving loan.
- There is a positive correlation between loan amount and installments. As the loan amount increases the installment also gets increases. The correlation value is 0.93.
- As the loan amount increases the interest rate is high.

# Inferences

- Term length: As the defaulters are more from term length 60, lending company has to identify the long term issues and modify the loan amount and interest rate accordingly
- Employment length: As the majority of defaulters are having 10 or more years of experience, the employment length is not a great factor to consider for the loan approval. Even with the employment length parameter other measures have to taken into consideration.
- Home ownership: Home ownership status as rented or mortgaged are the most likely to be defaulters.
- Verification status: As the verified customers are the most charged off the company has to implement strict verification process.
- Annual Income: Loan applicants who have annual income upto 40000 are the majority defaulters. Lending company should consider this for funding.
- Interest rate: High interest rate is main reason for defaulting. Company should review the interest rate.
- DTI: Many defaulters have high dti ration. This should be taken into consideration.

# Correlation Analysis

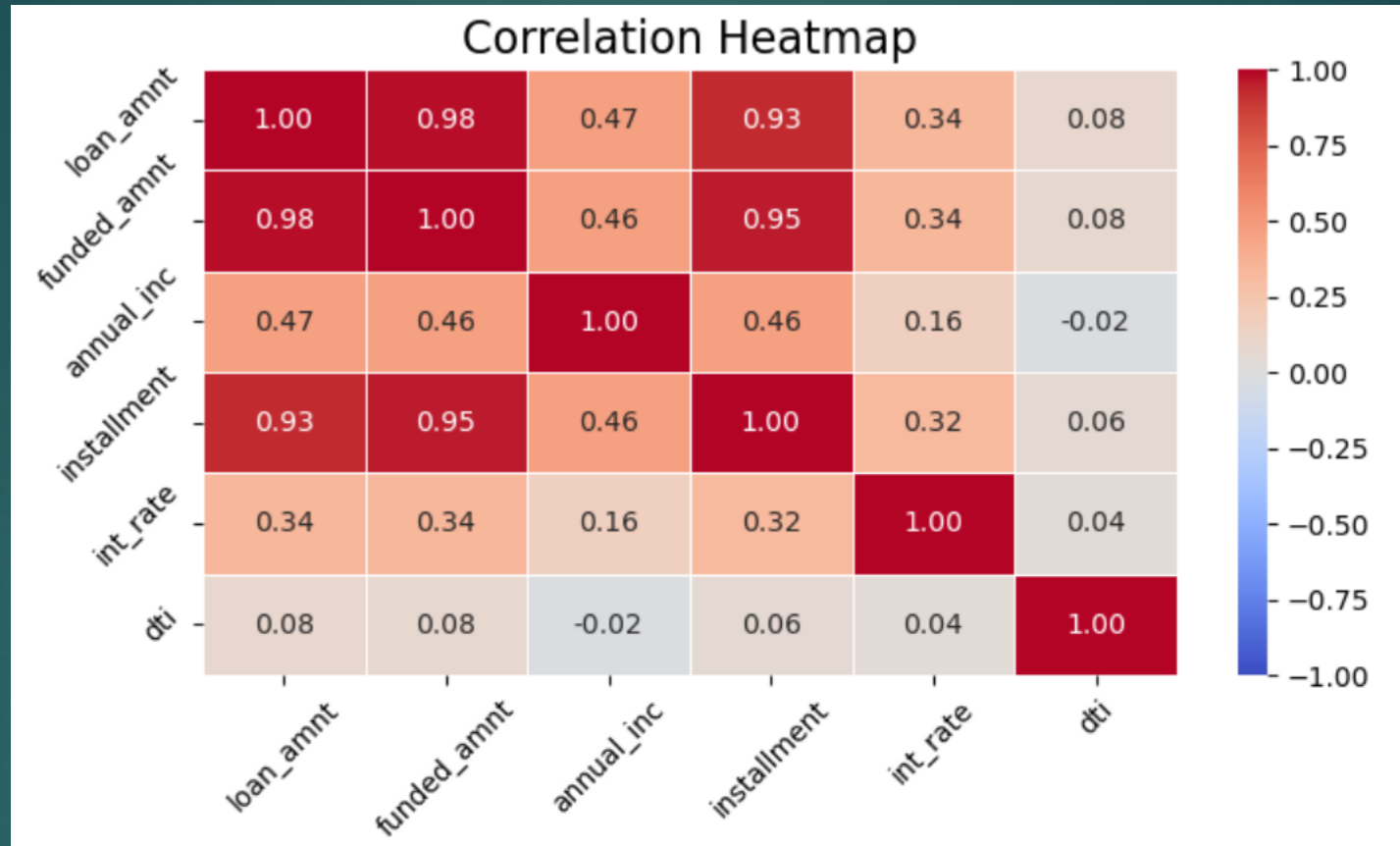
Correlation analysis is a statistical technique used to measure and describe the strength and direction of the relationship between two or more variables. It helps determine how one variable is related to another, indicating whether changes in one variable are associated with changes in another.

## Correlation Coefficient:

- The correlation coefficient is a number between **-1** and **+1** that indicates the direction and strength of the relationship between two variables.
- **+1** indicates a **perfect positive** linear relationship, meaning as one variable increases, the other variable also increases in perfect proportion.
- **-1** indicates a **perfect negative** linear relationship, meaning as one variable increases, the other decreases in perfect proportion.
- **0** indicates **no linear relationship** between the variables.



# Correlation Analysis



- There is a strong correlation between loan amount and instalment
- dti has weak correlation with other variables
- Annual income has negative correlation with dti

# Suggestions

- ▶ Verification Process Review: Evaluate the verification process to ensure the effectiveness to assess the creditworthiness of applicants.
- ▶ Consider home ownership: Take home ownership status into account during the process to determine the applicant's ability to repay the loan.
- ▶ Consider Income Levels : Annual income is an important parameter. Consider the income and employment duration to ensure loan affordability.
- ▶ Employment length consideration: High employment length is not a great parameter to determine the loan repay affordability. Analysis of this parameter along with others is important.
- ▶ Interest Rates Based on DTI Ratios: Evaluate the interest rate determination process and consider adjusting rates based on Debt-to-Income (DTI) ratios to better align with the borrower's repayment capacity.
- ▶ High Loan Amounts: Perform more comprehensive assessments for high loan amounts. Consider setting limits on loan amounts for higher-risk applicants to reduce the potential for defaults.
- ▶ Region: Analyze the general trend in different states and take required actions to limit the loan



Thank you