Lending Club Case Study

Submitter:

Vijay Kumar
Kushagra Bajpai

Our team

Vijay Kumar

Kushagra Bajpai

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. When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile.

Two types of **risks** are associated with the bank's decision:

1. If the applicant is likely to repay the loan, then not approving the loan results in a **loss of business** to the company
2. If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a **financial loss** for the company

- ✓ Like most other lending companies, 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.
- ✓ The primary objective of this case study is to look at existing loans data and identify the factors which contribute to charged off loans.

Objective

The objective of this exercise is to help Lending Club in reducing their financial risk when approving loan applications. Charged Off loans are the biggest contributor to these losses.
The company wants to understand the patterns in a loan application which can predict whether a borrower will be able to repay the loan or will they default.
We have been given access to existing loan application data with a wide range of information available about the loans and the borrowers. We need to use this data and help Lending Club make informed decisions when sanctioning risky loans.

Benefits we would achieve from this case study:

- ✓ Understanding key factors impacting loan default
- ✓ Isolate risky loans
- ✓ Strategies on increasing loan repayments
- ✓ Reduce Risk exposure
- ✓ Effectively manage portfolio

Dataset & Data Understanding

Provided data contains details of the passed loan applications, and whether these were loans were repaid or not.

The dataset contains

- Consumer Attributes
- Loan Attributes

Important Points:

- Only the loans which were approved are part of the dataset, loans which were rejected have not been included.
- The dataset collects information from new loans application and covers 111 Data points across various categories.

Data Cleaning and Analysis

Missing Values & Outlier Treatment

Steps Performed:

- **Load**ing the data using Pandas
- Remove columns with **NULL** values
- Remove columns with **Duplicate** values
- Remove columns with **high %age of missing** values
- **Fixing data types** of columns
- **Outlier Treatment**
- **Extracting additional columns from dates**

Univariate & Bi-Variate Analysis

Steps Performed:

- Performed Univariate Analysis
    - Ordered Categorical Variables Analysis
    - Ordered Categorical Variables Analysis
    - Quantative Variables Analysis
- Bi-variate Analysis
- Correlation Analysis

Univariate Analysis

Ordered Categorical Variable Analysis

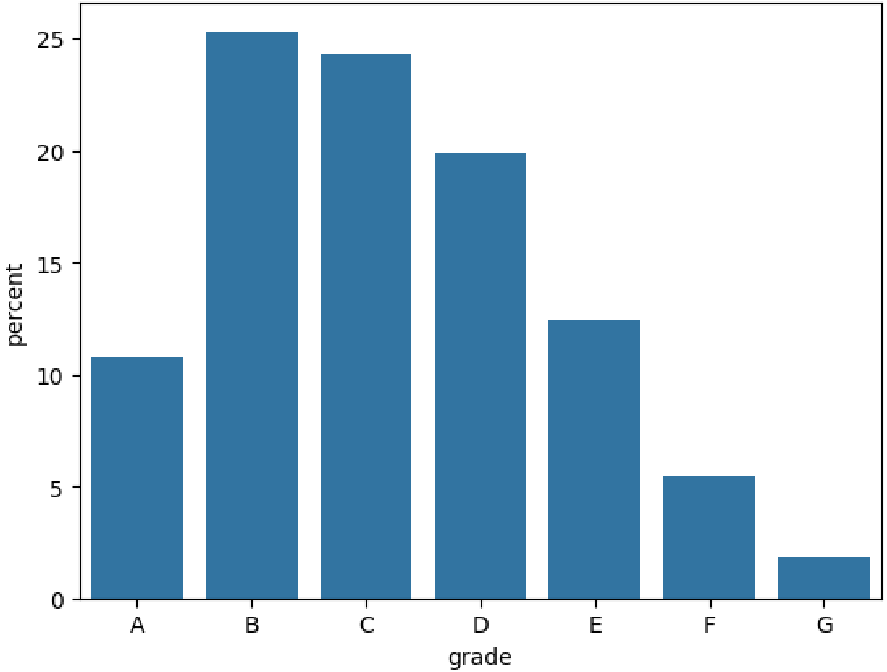
Below columns were identified for analysis:

Column Name	Meaning
Grade	Grade assigned to the loan account by lender
sub_grade	LC assigned loan subgrade
emp_length	Period of employment in years.
Term	The number of payments on the loan. Values are in months and can be either 36 or 60.
issue_d_y	- Year the loan was issued in.
issue_d_m	Month the loan was issued in
earliest_cr_line_m	The month of first reported credit. This can be used to calculate the credit age for the borrower. Generally speaking longer credit age tends to be more favorable
earliest_cr_line_y	The year of first reported credit. This can be used to calculate the credit age for the borrower. Generally speaking longer credit age tends to be more favorable

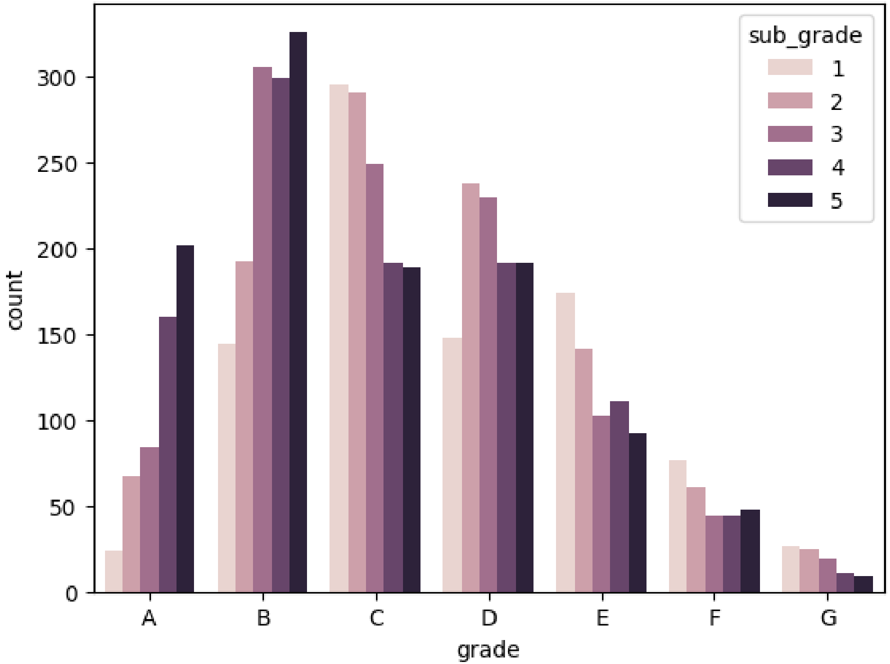
Analysis for Grades

Insight:

Most of the charged off loans are from category
- B with 25.31% defaults
- C with 24.30% defaults
- D with 19.94% defaults



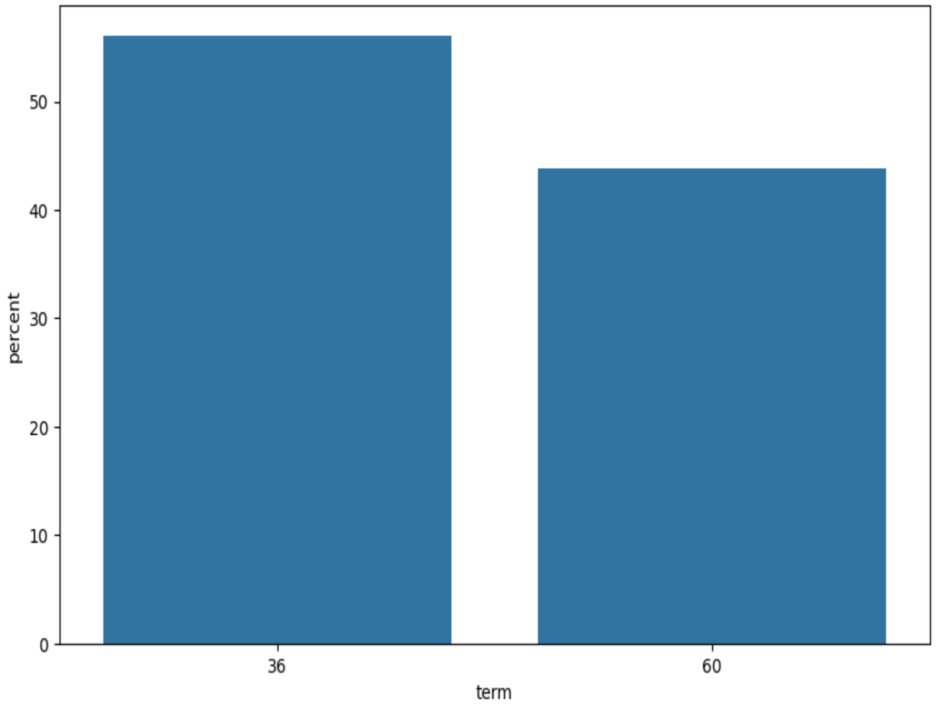
Analysis for Sub-Grades



Insight:

There isn't a clear pattern in charged off loans for sub-categories.
- Sub-Grade 5 is dominating in category B followed by 3 and 4
- Sub-Grade 1 has the most frequency in cat C followed by 2 and 3
- subgrade 2 is winning in cat D followed by 3 and 4.

Analysis for Loan Term



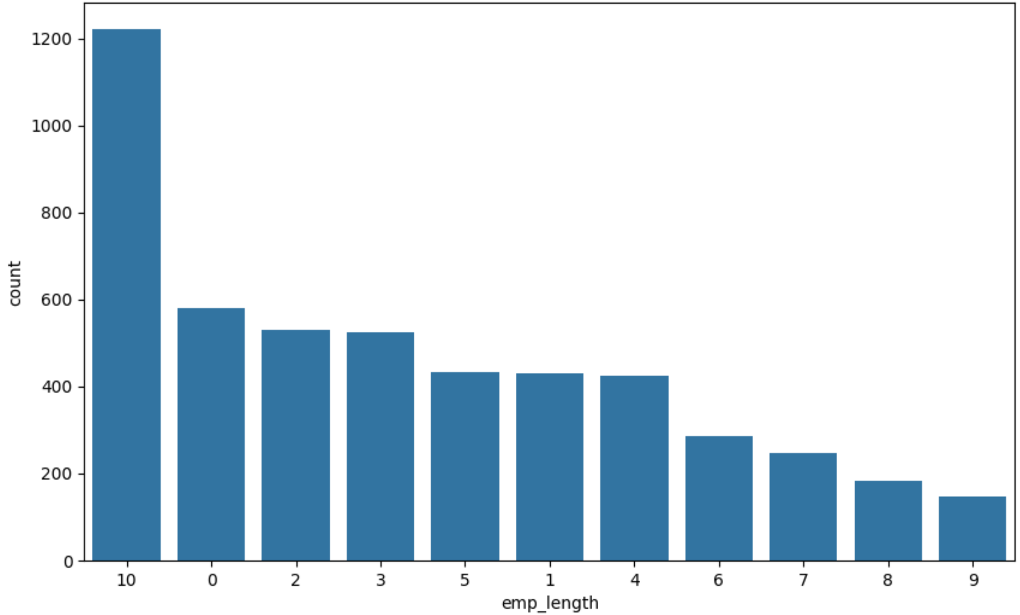
Insight:

- More than 50% of defaulted loans are taken for lower term.

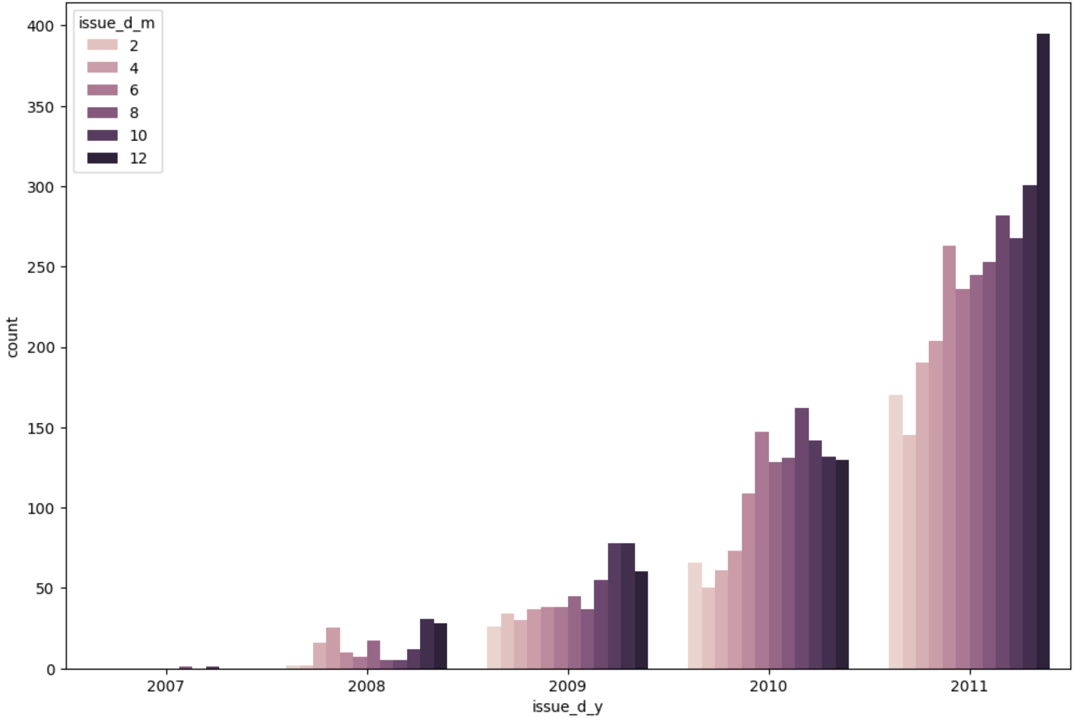
Analysis for Employee Length

Insight:

- Loan applications for applicants employed for more than 10years tends to be the most defaulted loans followed by <1 Year and 2 Year.



Analysis for Loan Year



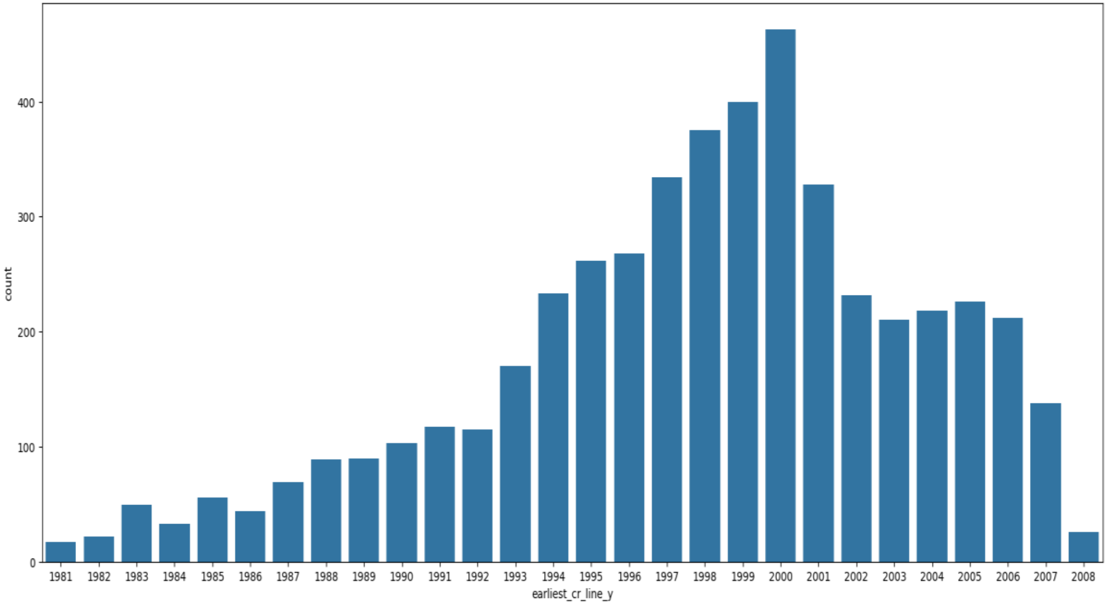
Insight:

- we could see an upward trend in number of defaulted loans over the years.
- most of the defaulted loans tends to be approved around end of year, this coincides with the holiday seasons.

Analysis for Earliest Credit Line

Insight:

- we could infer that a long credit history doesn't necessarily means the ability for repayment.
- Also charged off loans peaked for customers who started their credit in 2000 and is on a downward trend ever since.



Univariate Analysis

Un-Ordered Categorical Variable Analysis

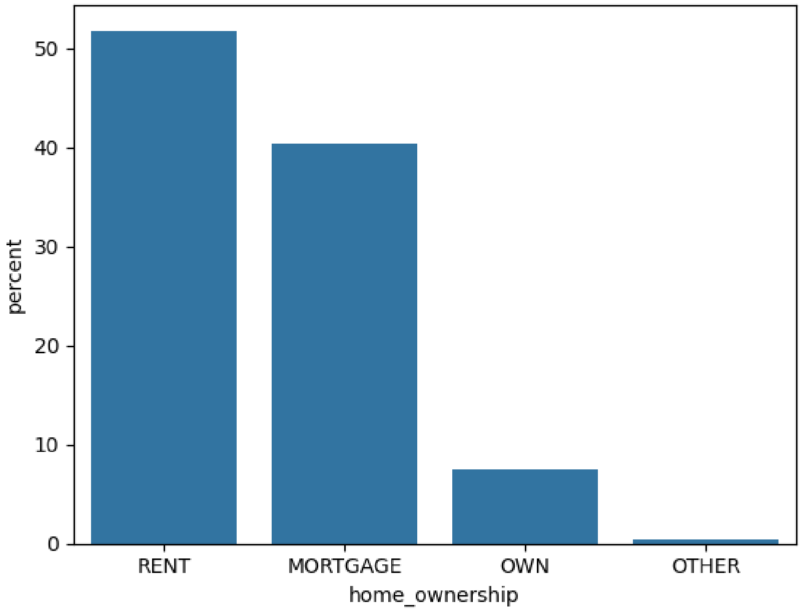
Below columns were identified for analysis:

Column Name	Meaning
home_ownership	Home Ownership status - RENT, OWN, MORTGAGE, OTHER
verification_status	Whether personal income was verified by lender
issue_d	Date the loan was issued in
purpose	category provided by the borrower for the loan request
zip_code	Area Zip Code
addr_state	2 Character state code provided by the customer when applying for the loan
delinq_2yrs	Number of incidences of delinquency in the borrower's credit file for the past 2 years
inq_last_6mths	Number of credit inquiries in past 6 Months
open_acc	The number of open credit lines in the borrower's credit file
pub_rec	Number of derogatory public records
total_acc	The total number of credit lines currently in the borrower's credit file
pub_rec_bankruptcies	Number of public record bankruptcies

Analysis for Home Ownership

Insight:

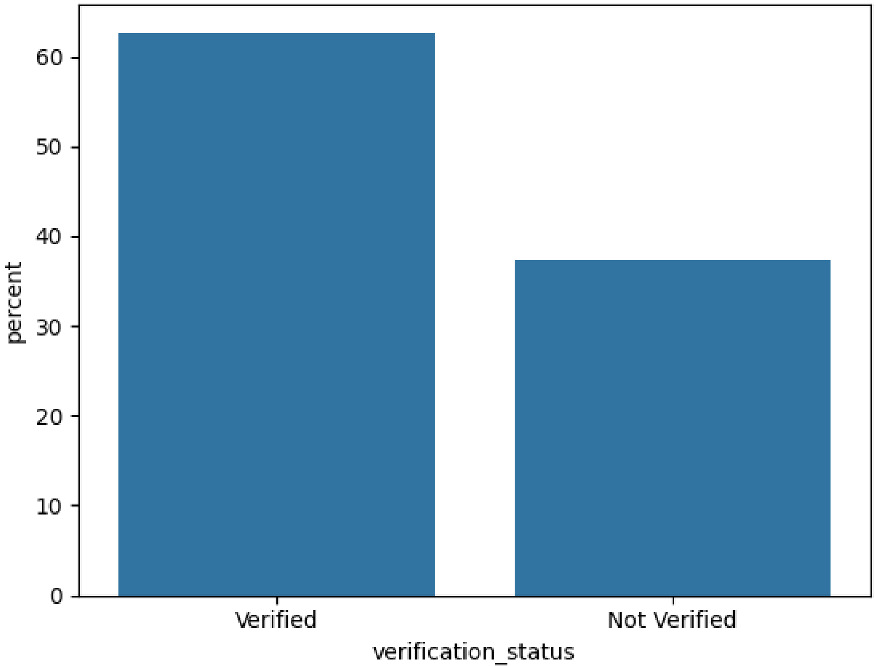
- Customers with Rented and Mortgaged homes make up for the majority of charged off loans.
  - This might be due to the additional financial commitments for rent and mortgage payments.
  - The LC should take extra precautions when considering borrowers ability to pay the installments if they have other fixed financial commitments.



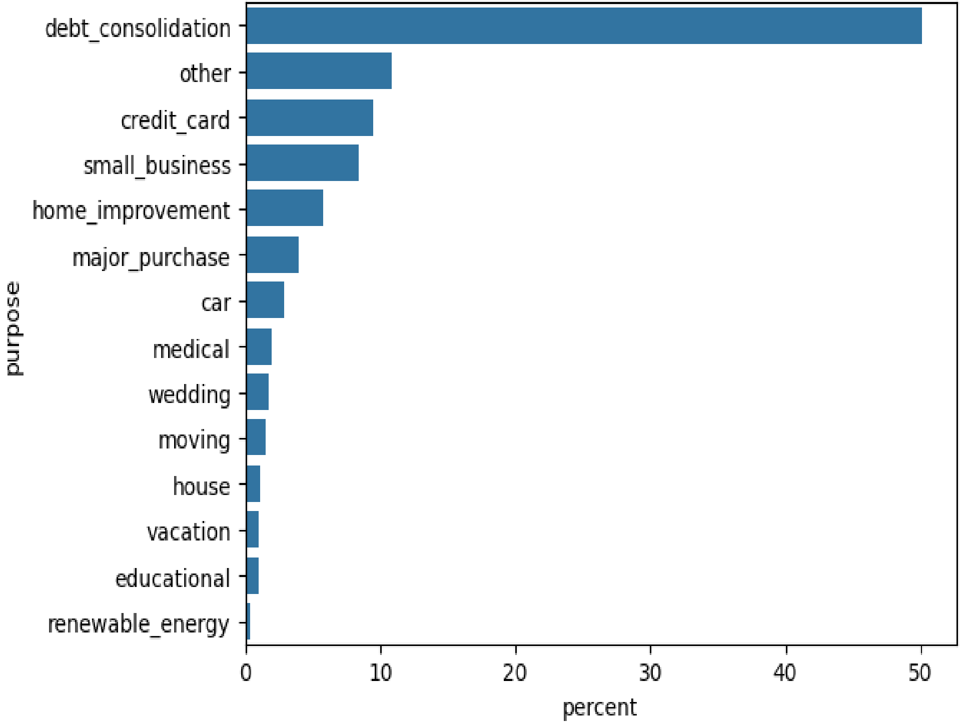
Analysis for Verification Status

Insight:

- Verified income is not a strong indicator of loan repayment capacity as most of the charged off loans are from verified category.



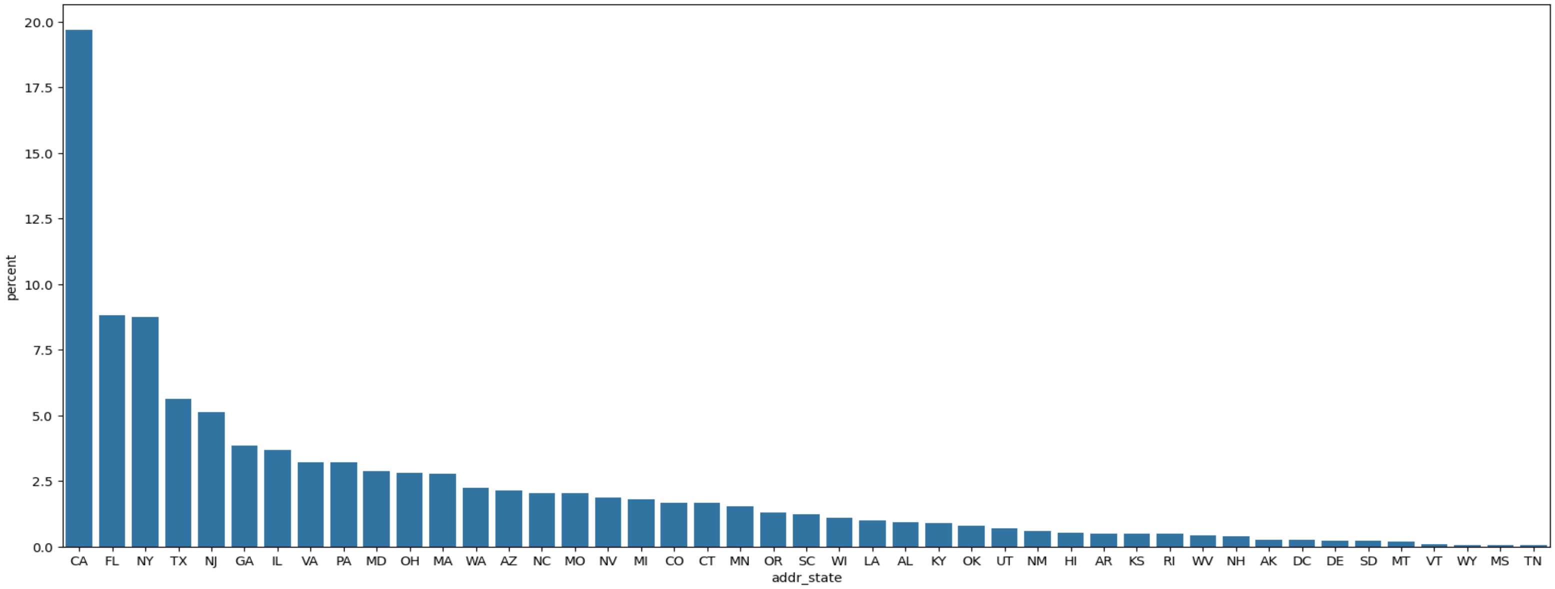
Analysis for Purpose



Insight:

- Customer who take loans for Debt Consolidation are at a higher risk of defaulting as they are already under financial pressure and might not meet the commitment.

Analysis for Address State

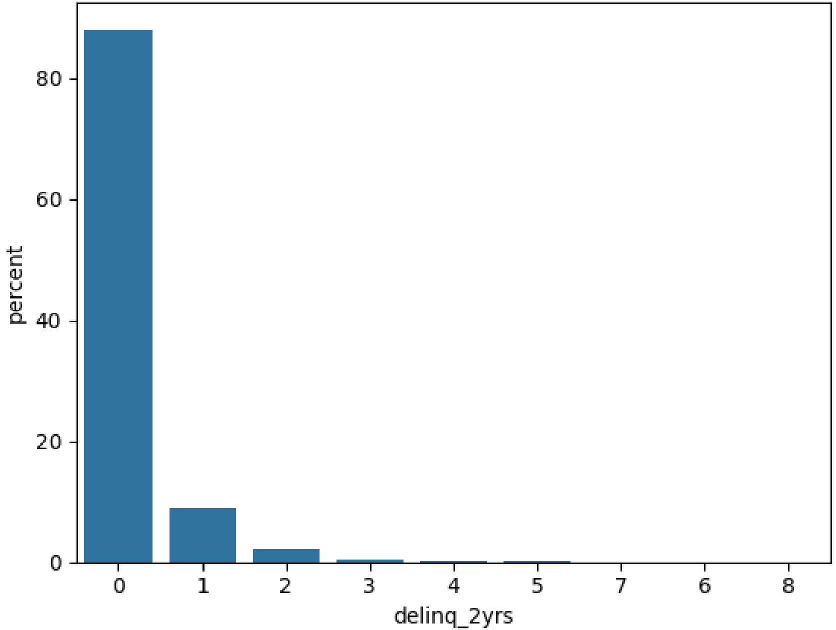


Insight:

- Most of the defaulted borrowers come from high GDP states such as California, Florida, New York, Texas and New Jersey.

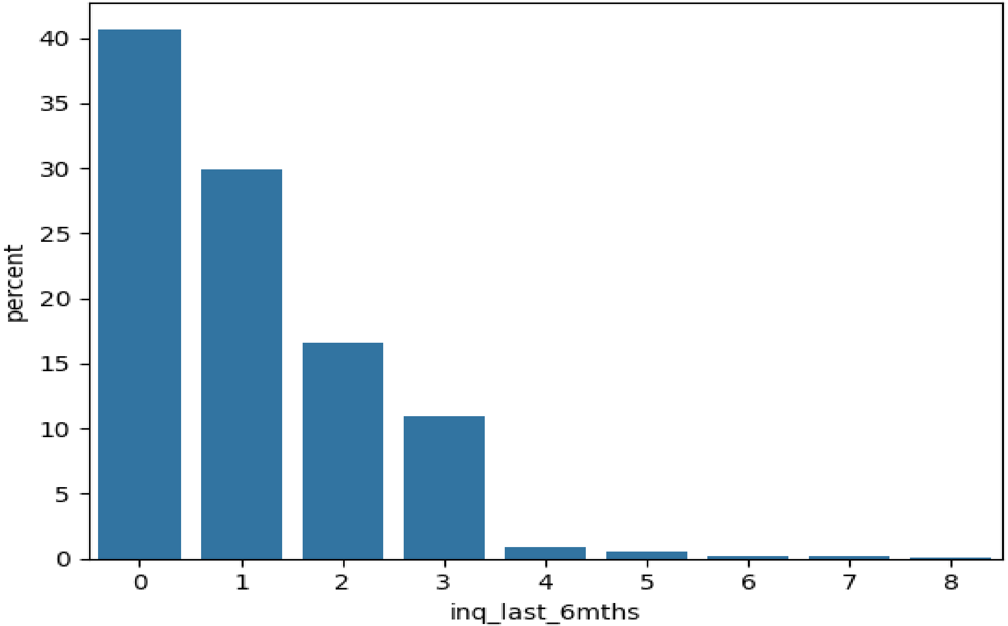
Analysis for Delinquency 2yrs

Insight:
- Customers with no past delinquency in 2 years have higher risk of defaulting.

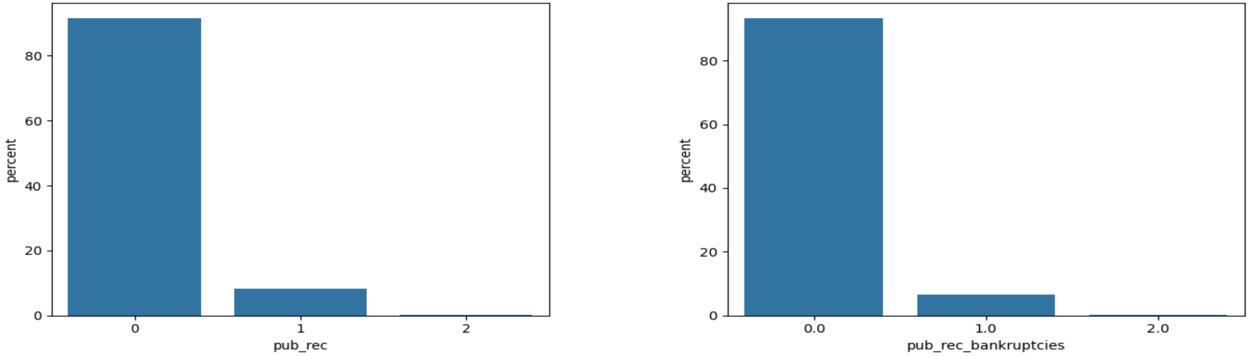


Analysis for Credit Inquiry in Last 6mths

Insight:
- surprisingly, credit hungry customers (With high number of credit inquiries) are less likely to default.



Analysis for public delinquency and bankruptcy records



Insight:

- Again surprisingly, customer with past public delinquency record or bankruptcy records are less likely to default on loans than the customer with no past public record.

 - This might be due to the fact the customers who have public record tends to be more diligent in paying back what they owe as they are aware of the negative impact of such loans.

Univariate Analysis

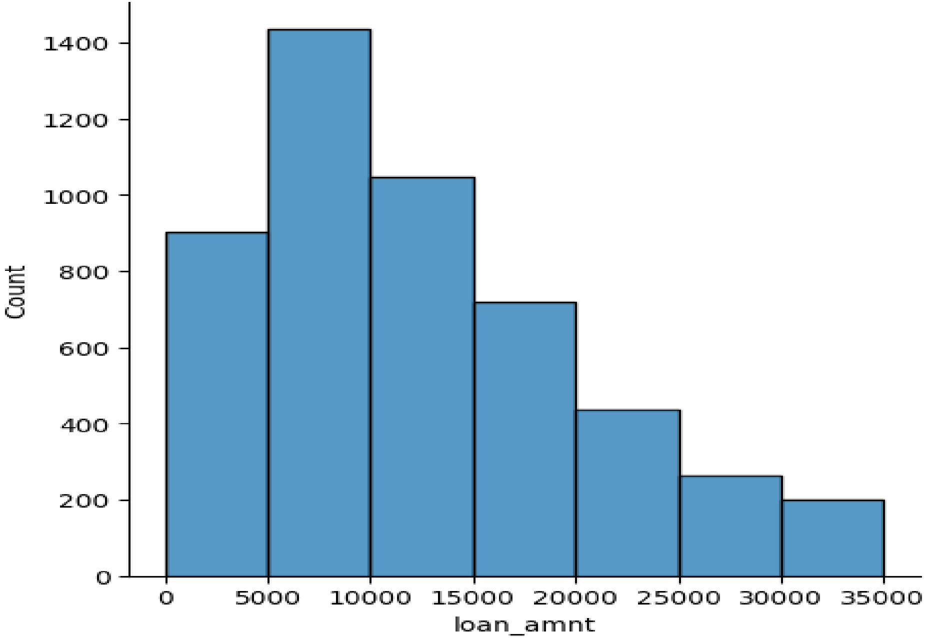
Quantative Variable Analysis

Below columns were identified for analysis:

Column Name	Meaning
loan_amnt	Amount of the loan applied for by the borrower
funded_amnt	Total amount committed to that loan at that point in time.
funded_amnt_inv	Total amount committed by investors for that loan at that point in time.
int_rate	Interest Rate
dti	Debt to Income ratio. DTI = (Current Monthly Loan payments - Mortgage - Requested Loan EMI)/reported monthly income
annual_inc	Customer reported annual income
installment	Loan Installment amount if approved.
revol_bal	Total credit revolving balance
revol_util	Revolving line utilization rate, or the amount of credit the borrower is using relative to all available revolving credit.

Analysis for Loan Amount

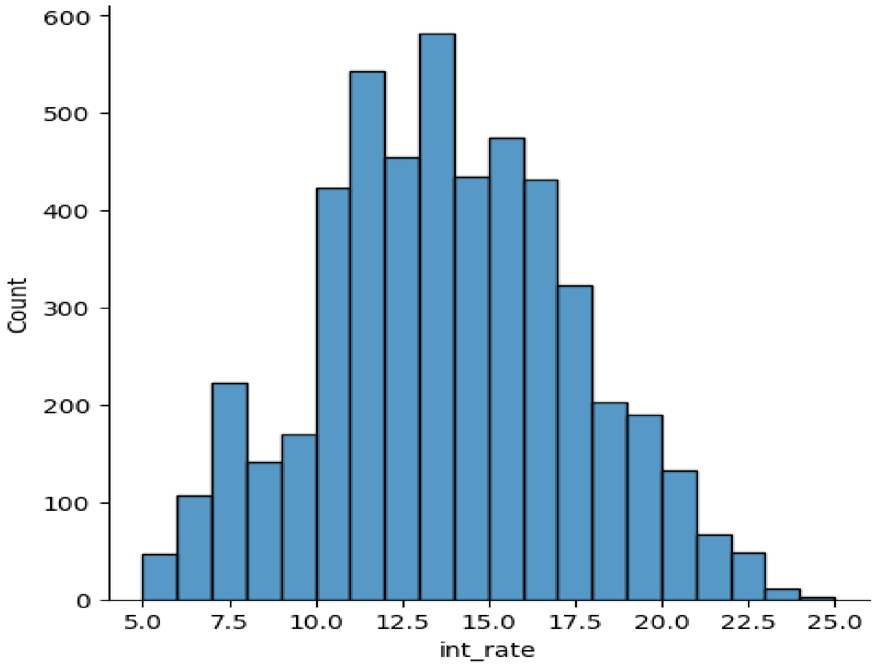
Insight:
- the loan amount for charged off loans is left skewed, meaning that borrowers who borrow smaller amounts in the range of 5-15K are at higher risk of defaulting.



Analysis for Interest Rate

Stats	Values
count	5001
mean	13.858008
std	3.631207
min	5.420000
25%	11.480000
50%	13.670000
75%	16.400000
max	24.400000

Insight:
- We see a sharp uptick in defaulted loans between interest rate 10 & 17 after which the trend seems to die down as int rate increases.

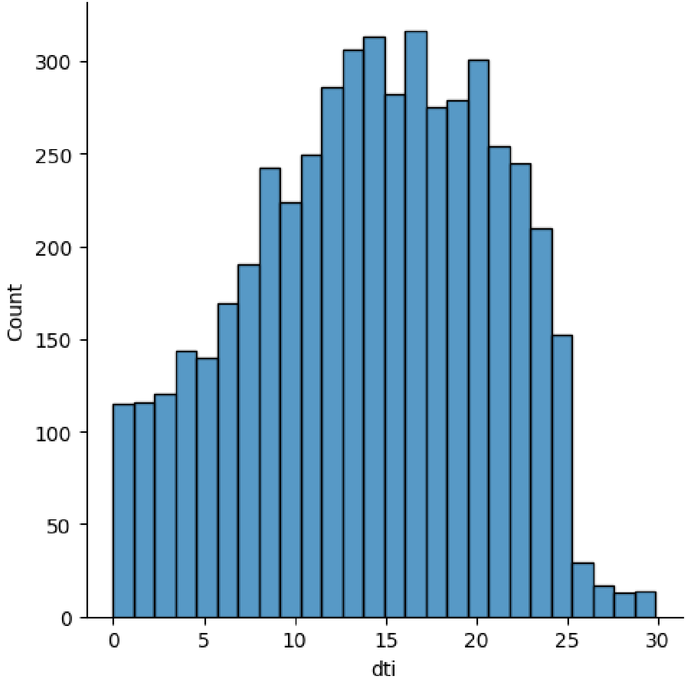


Analysis for Debt to Income(DTI) ratio

Stats	Values
count	5001
mean	14.144663
std	6.566810
min	0.000000
25%	9.250000
50%	14.470000
75%	19.400000
max	29.850000

Insight:

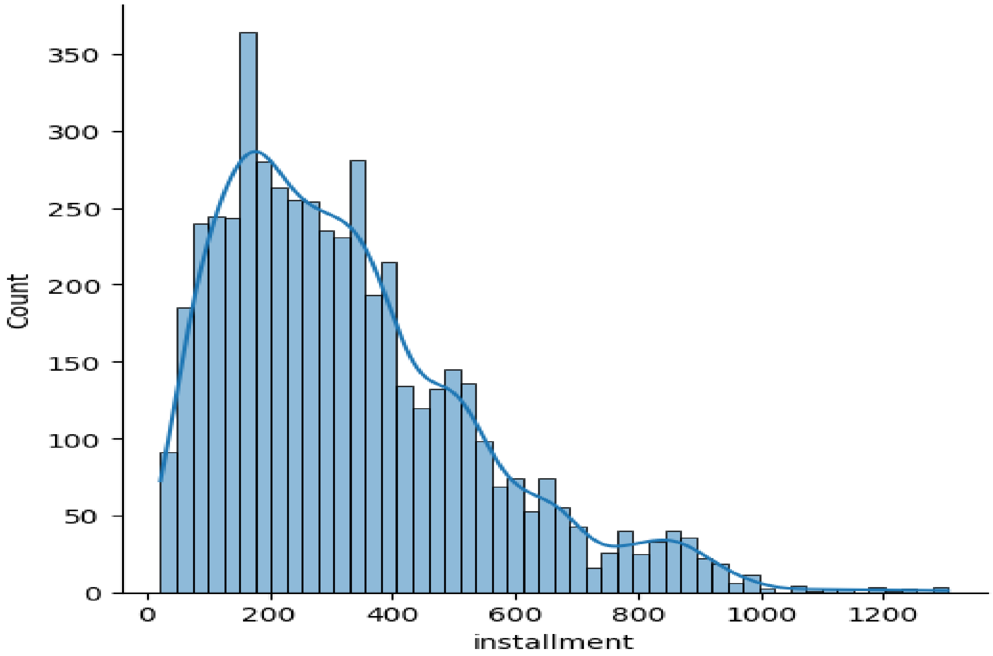
- majority of charged off loans have a DTI between 10-20%, counterintuitively number of charged off loans shows a downward trend for DTI higher than 20.



Analysis for Instalment

Insight:

- From this we see that loans with installments between 75-300 faced issues with repayment



Bi-variate Analysis

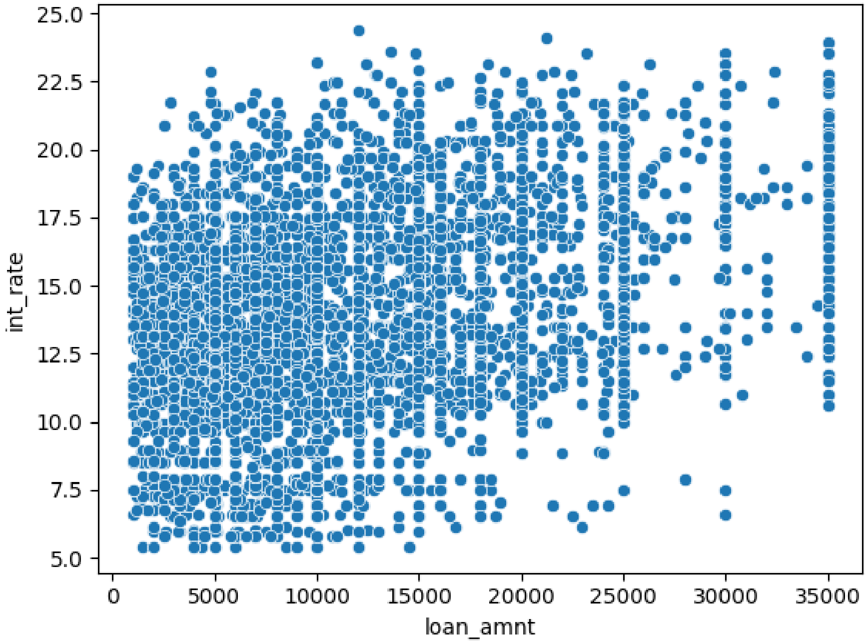
Defining variable pairs

Below pairs were identified for analysis:

Paris	Meaning
loan_amnt and int_rate	Loan Amount vs. Interest Rate
grade and loan_amnt with subgrade	Grade vs. Loan Amount(With Sub-grade)
term and loan_amnt	Loan Term vs. Loan Amount
emp_length and loan_amnt	Employment Tenure vs. Loan Amount
annual_inc and loan_amount	Annual Income vs. Loan Amount
loan_amnt and term	Loan Amount vs. Loan Term
home_ownership and loan_amnt	House Ownership vs. Loan Amount
verification_status and loan_amnt	Verification Status vs. Loan Amount

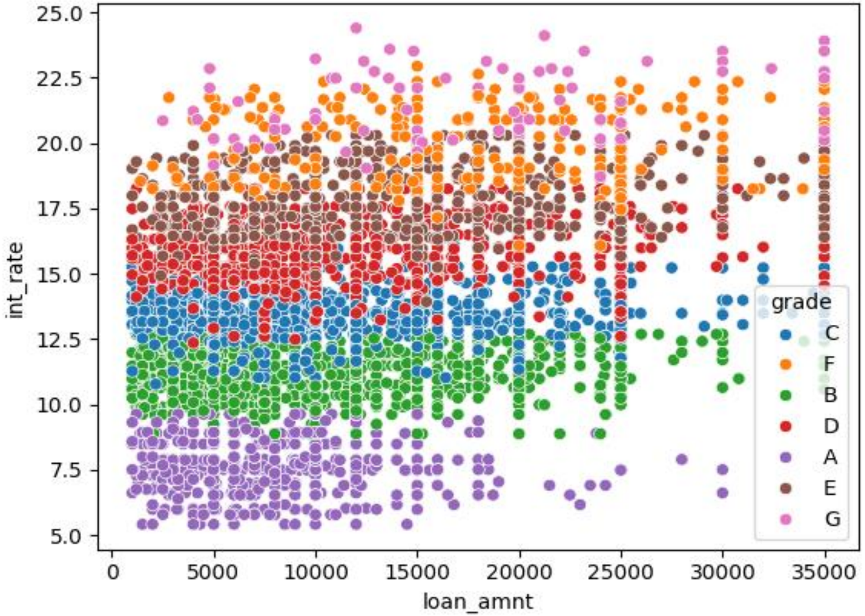
Bi-variate Analysis for Loan Amount and Interest Rate

Insight:
- simply looking at the loan_amnt and int_rate scatter doesn't tell us anything
- loans look to be concentrated on the left side which is known from past analysis.
- let's add hues and see how that changes



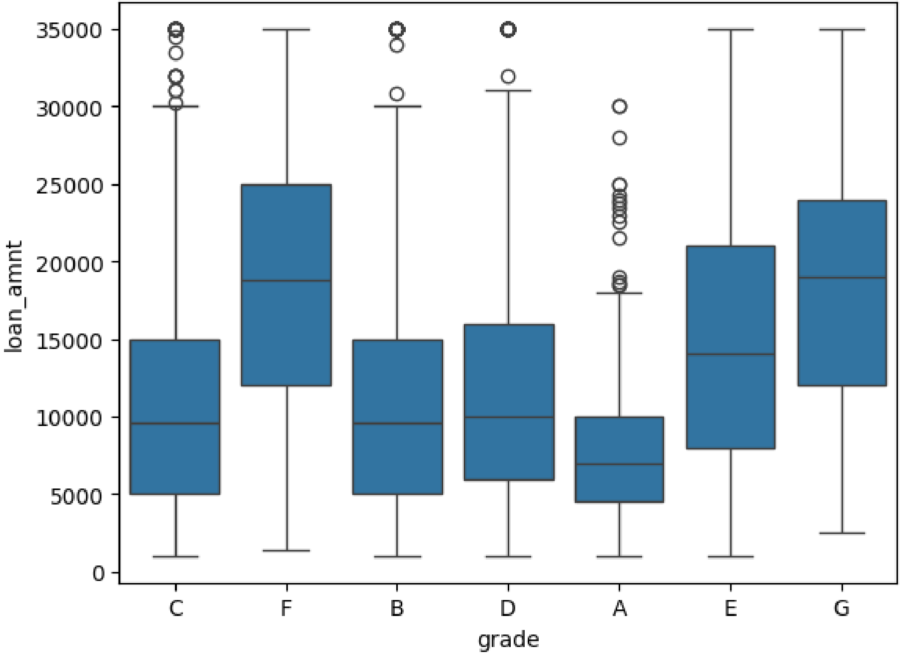
Bi-variate Analysis for Loan Amount and Interest Rate Cont.

Insight:
- From this we can see that Grade A, B and C loans were given at a lower int rate than D,E,F & G.
- there is some overlap but that's understandable



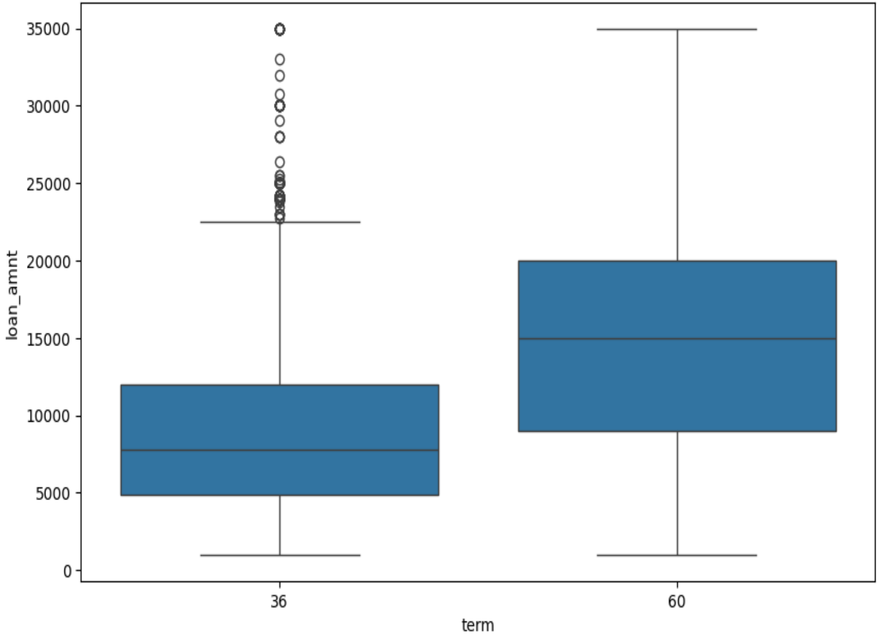
Bi-variate Analysis for Grade & Loan Amount

Insight:
- barring a few outliers, lower grade loans F,D,E and G have higher max and median values than higher grade loans A,B and C.
- This means that higher loan amounts were sanctioned for lower grade loans at a higher interest rate as inferred from last plot. increasing the risk for LC.

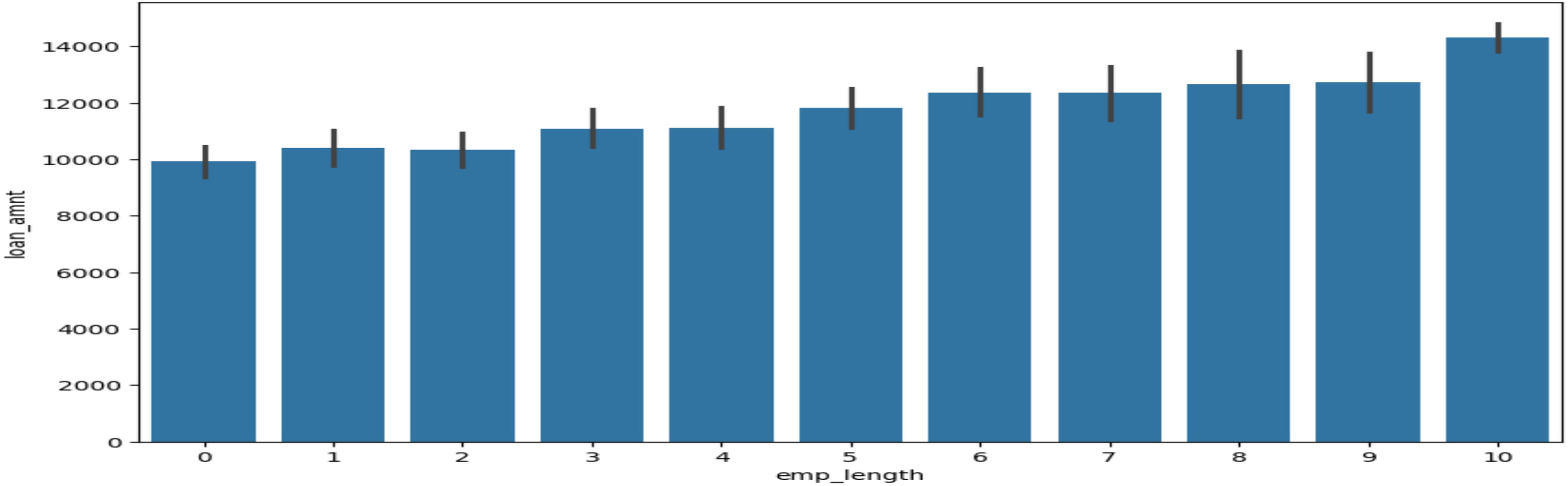


Bi-variate Analysis for Loan Amount & Term

Insight:
- lower term loans tend to be for lower amount with certain outlier.

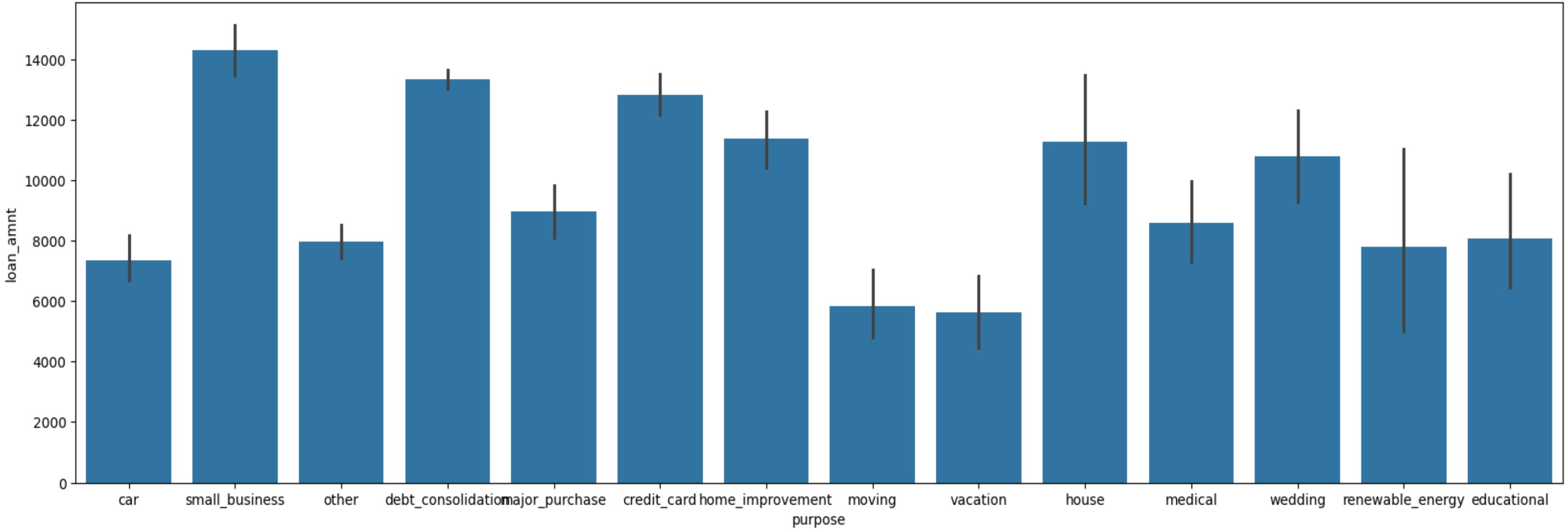


Bi-variate Analysis for Employee Length & Loan Amount



Insight:
we see an upward trend in loan amount as the borrower employment length grows.
- lets see what they are spending the loan amount on.

Bi-variate Analysis for Purpose & Loan Amount

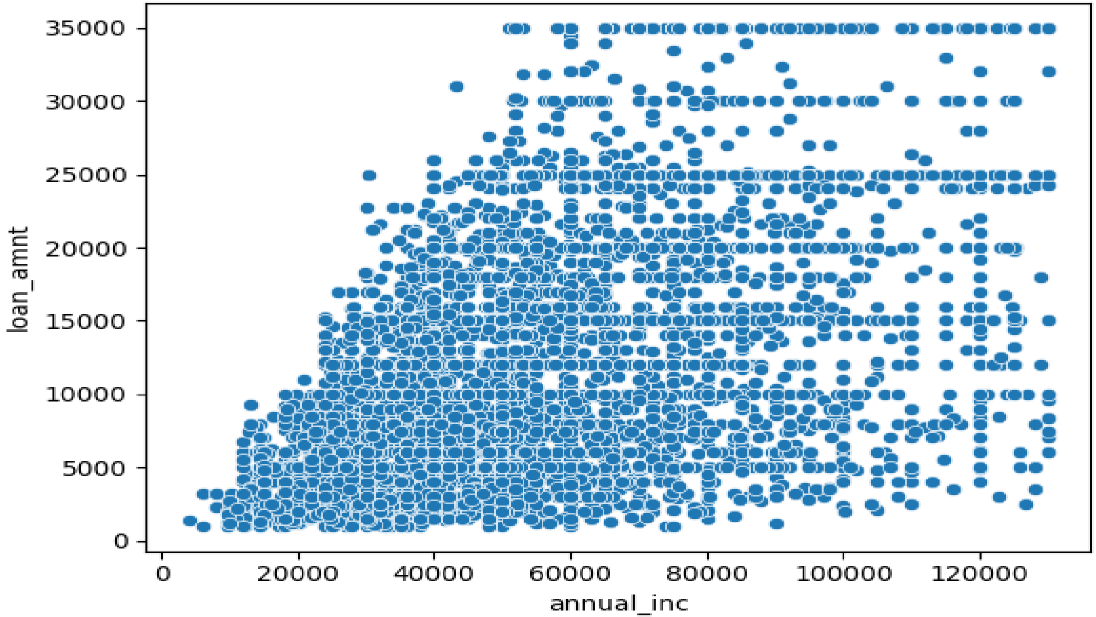


Insight:

- Most high value loans are sanctioned for Small Business, followed by debt consolidation, house, home_improvement and wedding

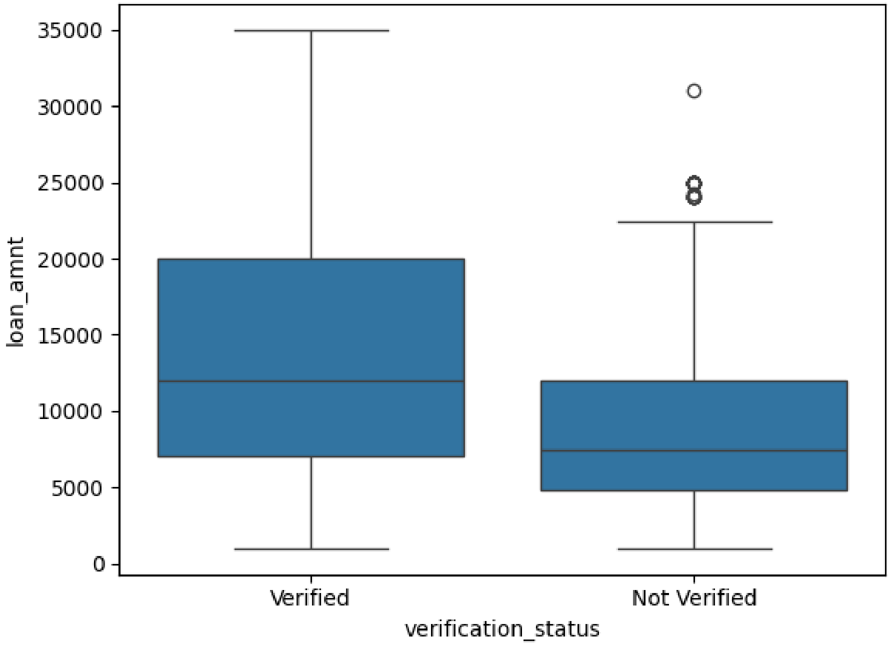
Bi-variate Analysis for Annual Income & Loan Amount

Insight:
- most of the loans are concentrated in the lower left and gradually scatter to up-right.
- most defaulted loans are taken for smaller amounts and by people with lower income.



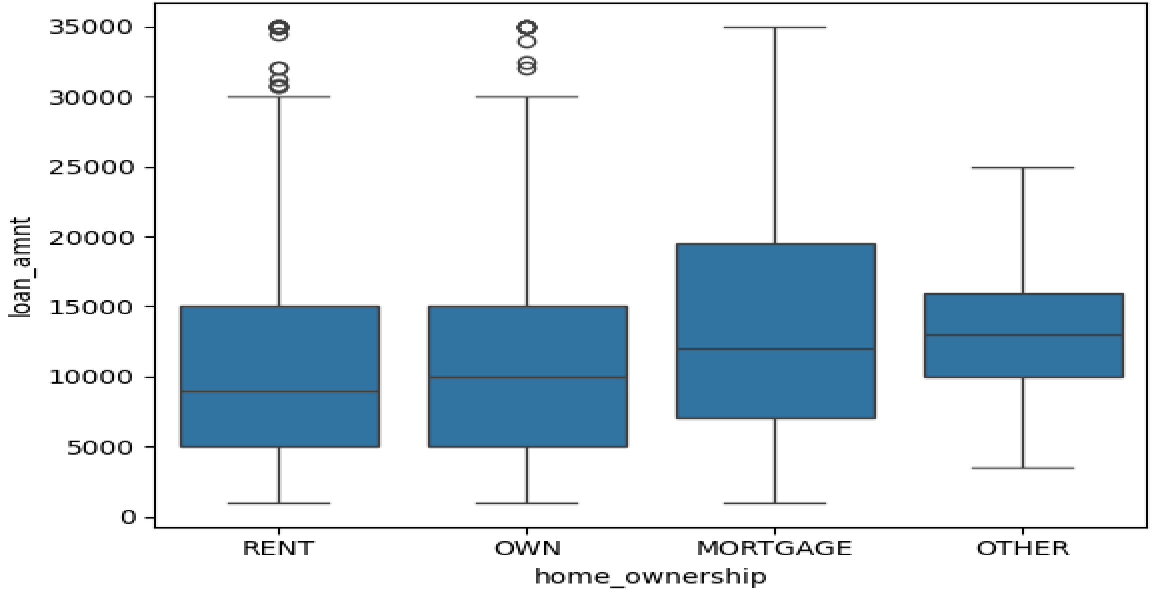
Bi-variate Analysis for Verification Status & Loan Amount

Insight:
- higher loan amounts are sanctioned for verified income status.
- there are some outlier in the not verified but looks like smaller loan amount are either not verified or rejected for which data is not available.

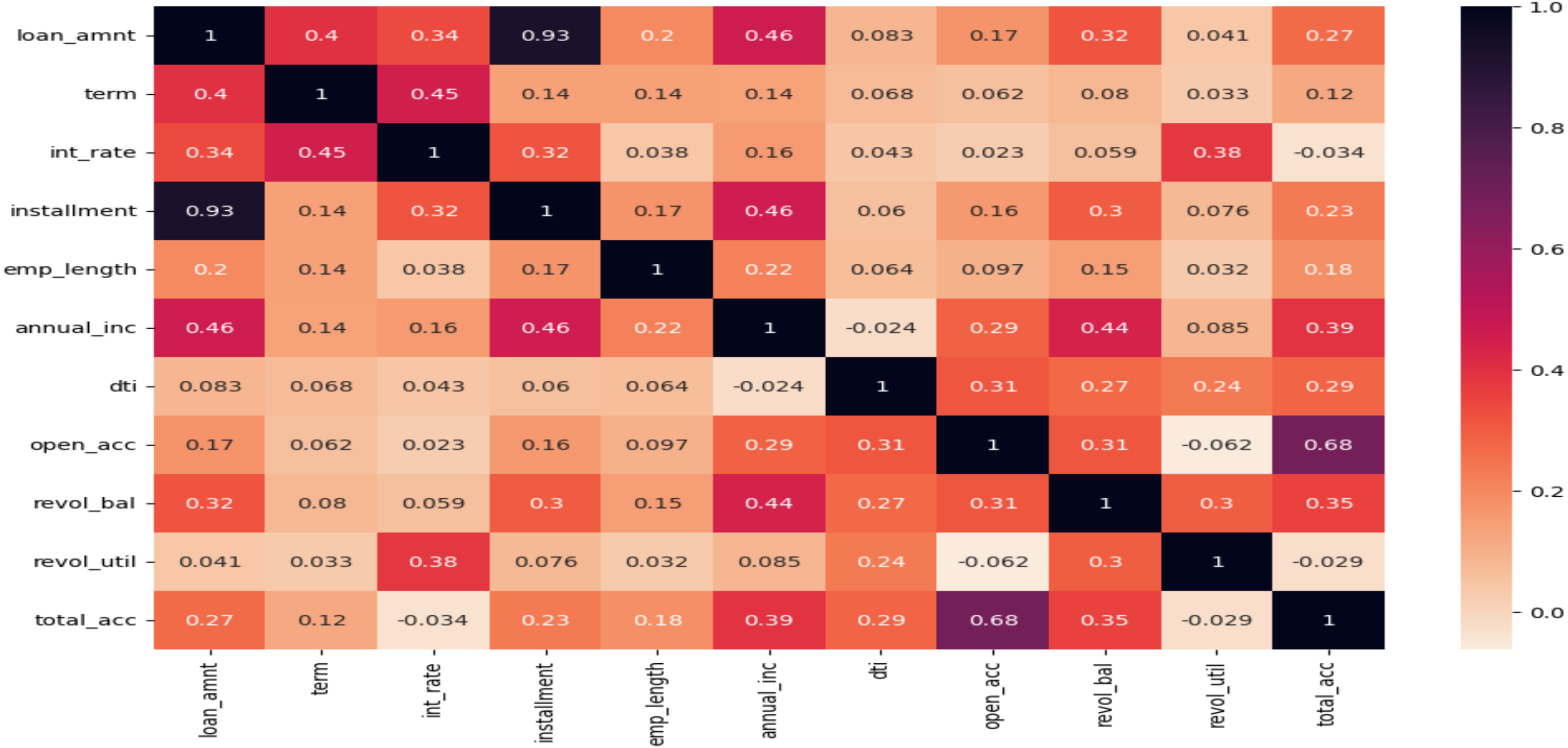


Bi-variate Analysis for Home Ownership & Loan Amount

Insight:
- people who rent or own houses have similar loan requirement which people who have mortgages tend to go for higher loan amounts which could be explained by their higher financial responsibilities.



Correlation Analysis



Correlation Analysis Contd.

Very Strong Correlations (.8 and above)	Strong Correlations (.6-.79)	Moderate Correlations (0.4-0.59)	Weak Correlations (0.2 - 0.39)	Very Weak Correlations (0-0.19)
• loan amount with installment	• open accounts with total accounts	• term with loan_amnt • term with annual_inc • term with int_rate • installment with annual_inc • annual_inc with revol_balance	• loan_amnt with emp_length_mapping, int_rate and total_acc • int_rate with installment and revol_util • installment with revol_bal and total_acc • emp_length with annual_inc • annual_inc with open_acc and total_acc • dti with open_acc, revol_bal, revol_util and total_acc • open_acc with revol_bal • revol_bal with revol_util and total_acc	• loan_amnt with dti, revol_util and open_acc • term with installment, emp_length, annual_inc, dti,open_acc, revol_bal,revol_util & total_acc • int_rate with emp_length, annual_inc,dti, open_acc, revol_bal, total_acc • installments with emp_length, dti, open_acc, revol_util • emp_length with dti, open_acc, revol_bal, revol_util, total_acc • annual_inc with dti, revol_util • open_acc with revol_util • revol_util with total_acc

Analysis Summary & Recommendation

Loan Grade

- Grade B, C and D have the largest contribution in defaulted loans. LC needs better guidelines and assessments for grading the loans.
- There isn't a clear pattern for sub-grades. LC needs a better framework to categories the loans.

Employment Experience

- Loans given to borrowers with 10+ Years of expirience are at a higher risk of being charged off.
- There is a big jump (3%-24%) in loan defaults from 9 to 10+ Years, suggesting that granular details are needed for 10+ years expirience category. Currently 10+ holds everyone with more than 9 years of experience.

Loan Term

- More than 50% of defaulted loans are taken for lower term, LC should consider the term decisions as small term loans tend to have higher installments which can impact repayment capacity.

Analysis Summary & Recommendation Cont.

Trend

- There is an upward trend in loan defaults since 2007 with 2011 seeing maximum number of defaulted loans.
- Most of the defaulted loans are approved around end of year, this co-incides with the holiday seasons.

Credit History

- Long credit history doesn't necesarily means the ability or repayment.

Home Ownership

- Customers with Rented and Mortgaged homes make up for 90% of charged off loans. LC need to ensure the borrower with Rented o Mortgaged accomodations have enough surplus to meet their EMI requirement.

Income Verification

- Verified income loans makes up for over 60% of defaulted loans. LC need to ensure that income verification process is more robust and take other factors into account such as DTI ratio, existing commitments etc.

Analysis Summary & Recommendation

Purpose

- Customer who take loans for debt_consolidation are at a higher risk of defaulting as they are already under financial pressure and might not meet the commitment.

Geography

- Most of the defaulted borrowers come from high GDP states such as California, Florida, New York, Texas and New Jersey. LC needs to take local trends into account when sanctioning loans.

Past Credit History

- Surprisingly, credit hungry customers (With high number of credit inquiries) are less likely to default.
- Customers with no past public deliquency record or bankruptcy recoreds are 8X more likely to default on loans than the customer with past public records.

Loan Amount

- Loans in the range of 5-15K are at higher risk of defaulting. LC can cover their risks by capping the loan amount or by increasing interest.

Analysis Summary & Recommendation

Interest Rate

- Maximum defaulted loans have int_rate between 10 and 17 percent.
- Surprisingly, loans with higher interest rate >18% have very low default rates. LC can use this to increase int_rate for high risk loans.

Annual Income

- Majority of charged off loans are from income range 37K-70K.

Installments

- Loans with installments between 75-300 faced issues with repayment. LC can offer longer terms for these loans to reduce installment burden.

Thank you!