LENDING CLUB CASE STUDY

INDRAMANI PANDEY & SANJIT M M

PROBLEM STATEMENT

You work for a **consumer finance company** which specializes in lending various types of loans to urban customers. 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:

- If the applicant is **likely to repay the loan**, then not approving the loan results in a **loss of business** to the company
- 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

AIM

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.

(Minimize the risk of loss while lending money)

KEY STEPS

- Data Sourcing
- Data Cleaning
- Univariate Analysis
- Bivariate Analysis
- Multivariate Analysis

DATA SOURCING

Data source – A loan dataset file named **‘loan.csv’**

There is also another file called **‘Data_Dictionary.csv’** which gives information about every column of the loan dataset

DATA CLEANING

- Null value check
  - Dropping columns which has null values > 40%
- Dropping unwanted columns (columns not required for the analysis)
- Ignoring all ‘current’ loans
- Handling missing values
  - Option 1- Dropping the values if the count is less (‘revol_util’)
  - Option 2- Filling the values with median or mode (‘emp_length’ and ‘pub_rec_bankruptcies’)
- Numerical and categorical columns
- Standardizing data
- Outliers (‘annual_inc’)

UNIVARIATE ANALYSIS

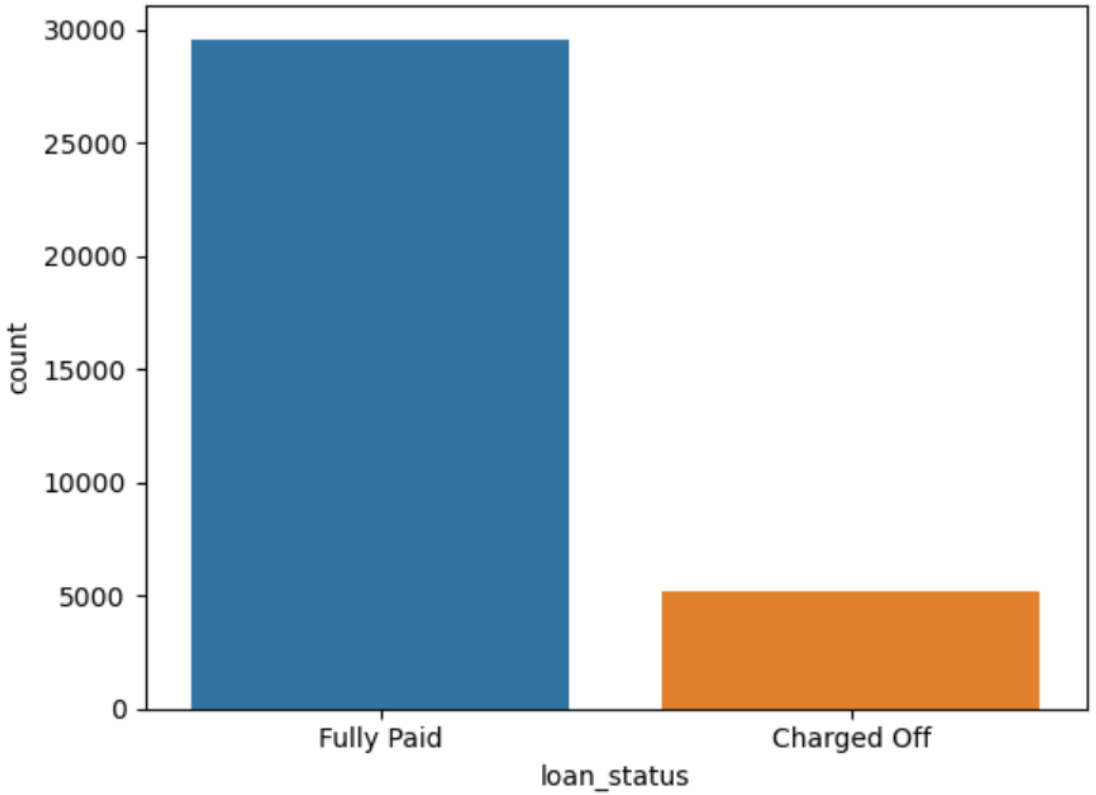
1) Categorical univariate analysis

2) Numerical univariate analysis

3) Segmented univariate analysis

CATEGORICAL UNIVARIATE ANALYSIS

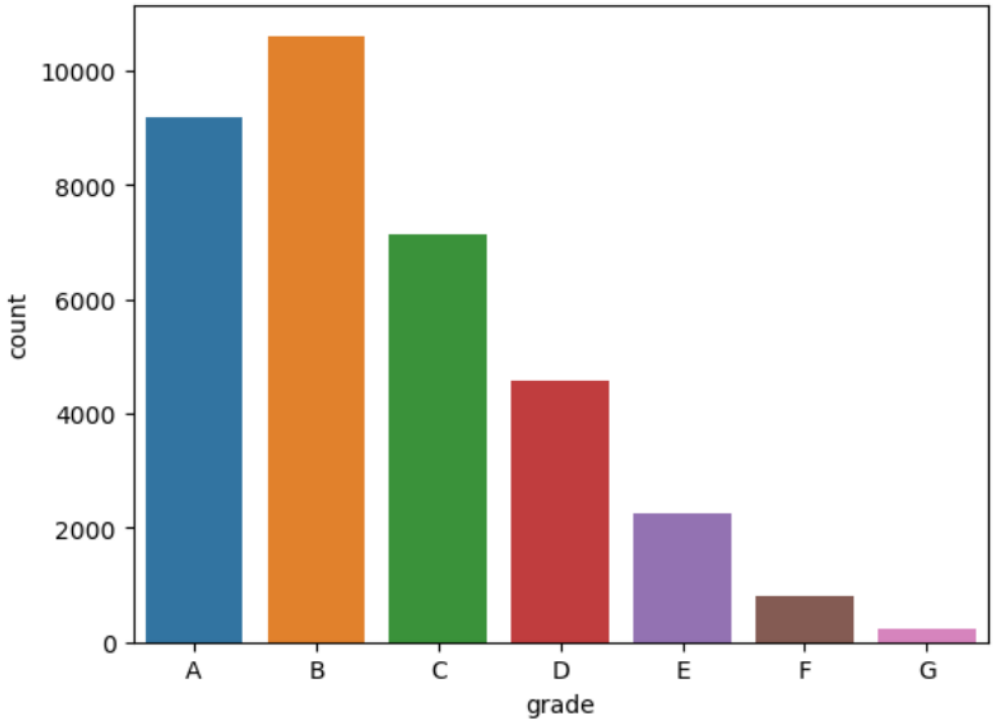
Analyzing the 'loan_status' values



It is observed that 14.78% of the customers are defaulters

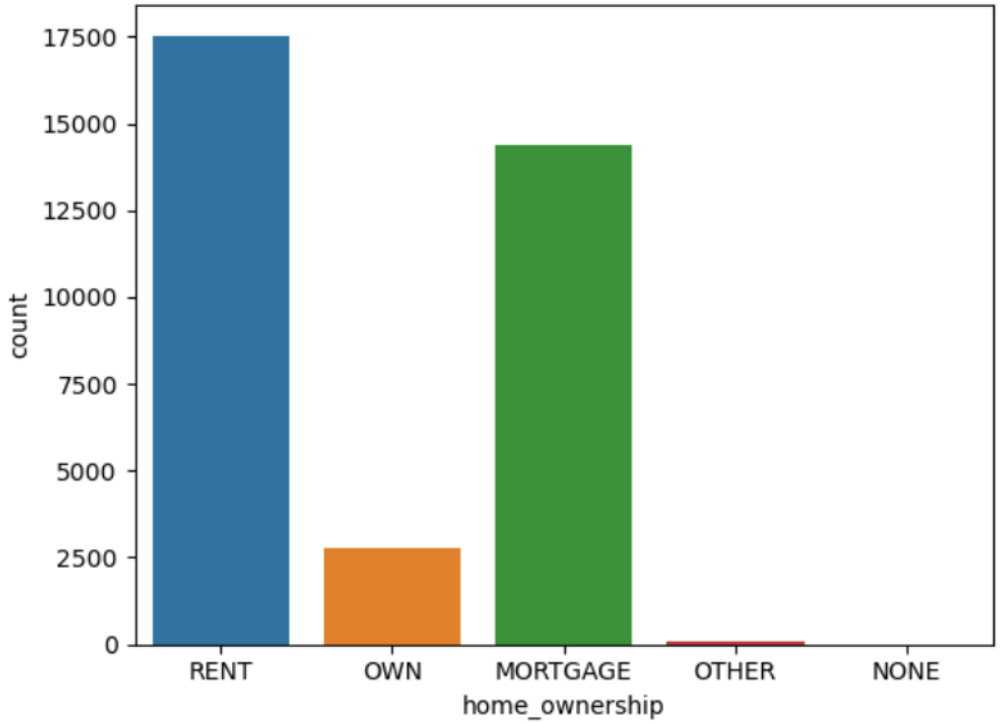
CATEGORICAL UNIVARIATE ANALYSIS

Analyzing the 'grade' values



It is observed that as grade increases, no of customers decreases. Most of the customers belong to the grade A, B and C (B is the maximum)

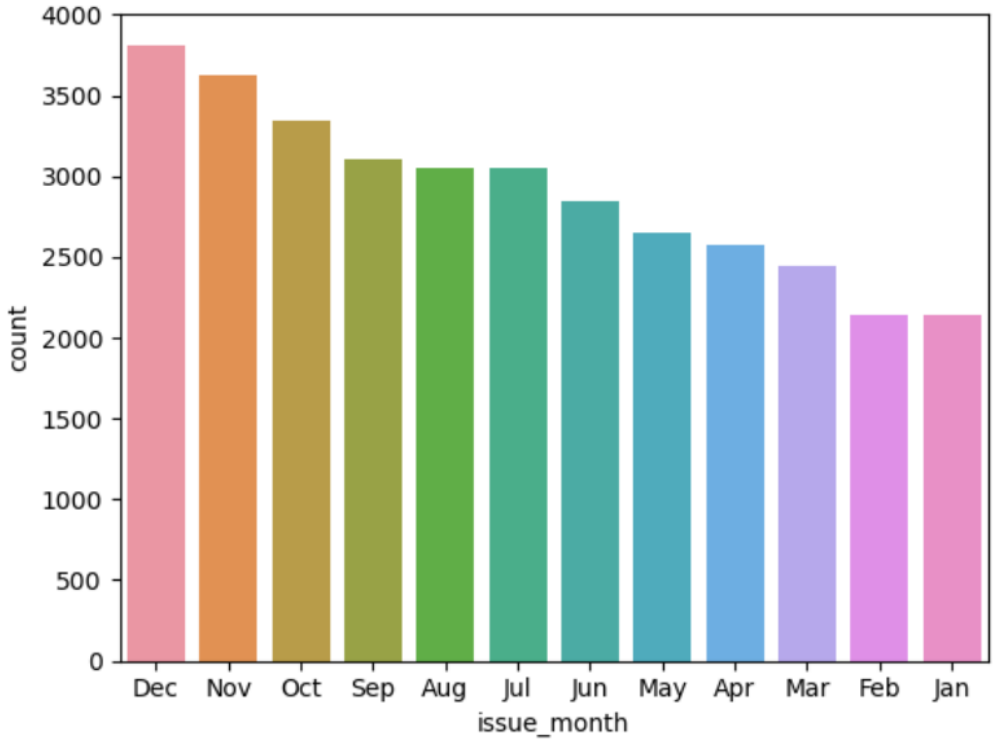
Analyzing the 'home_ownership' values



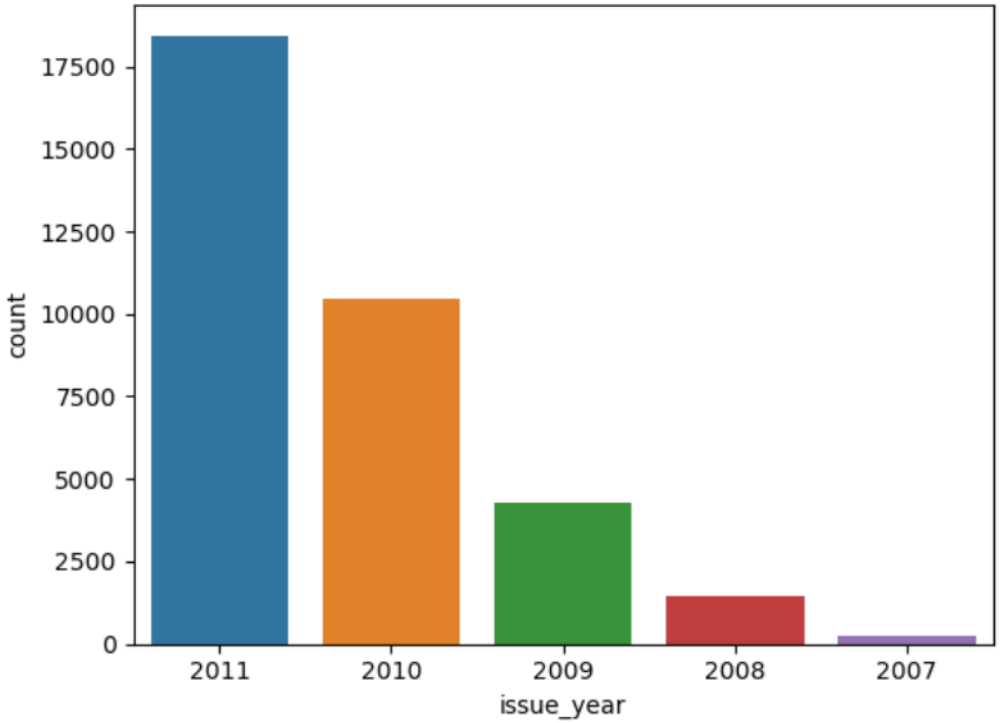
Majority of the customers live in either a rented or a mortgaged house

CATEGORICAL UNIVARIATE ANALYSIS

Analyzing 'issue_month’



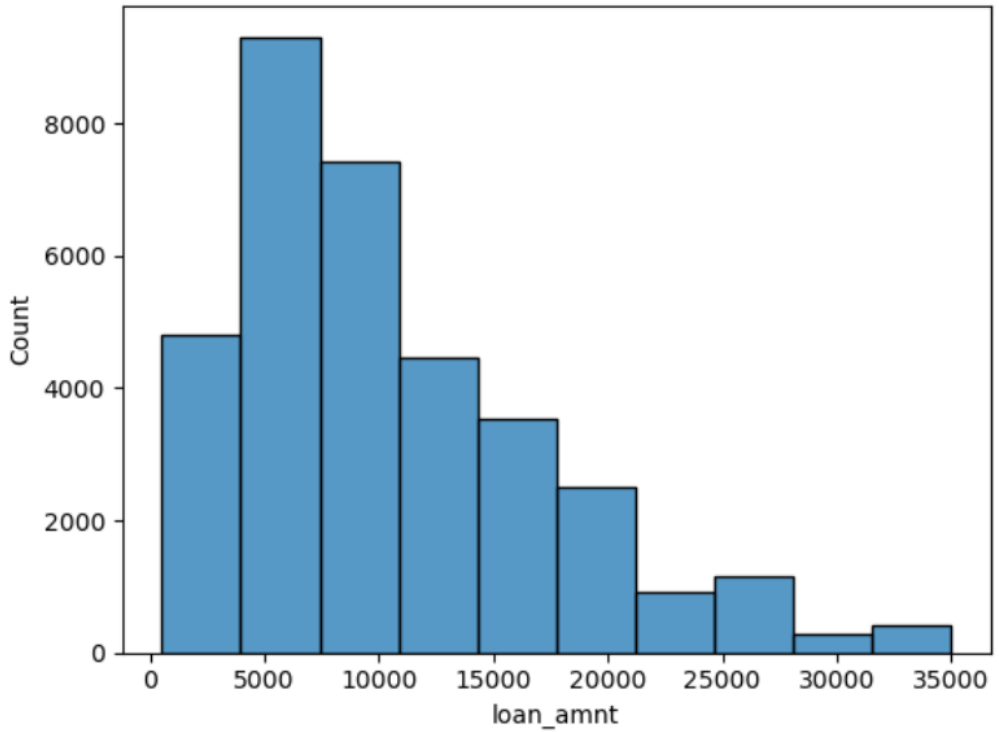
Analyzing 'issue_year'



The above 2 plots show that there were more customers when the loan was issued in the year 2011 and especially in the month December

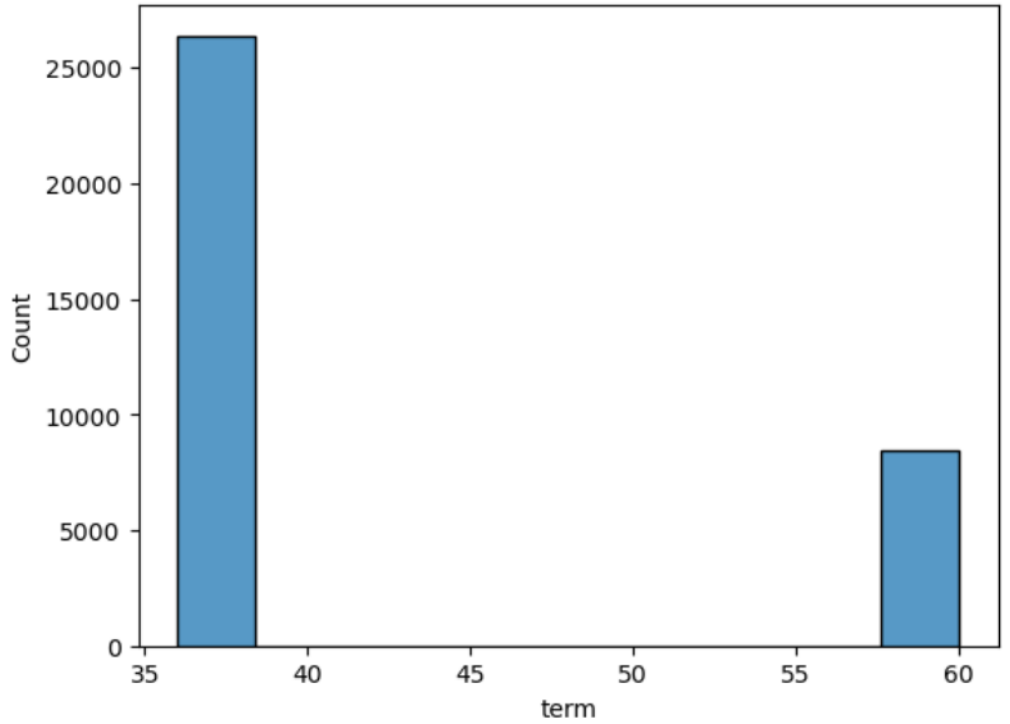
NUMERICAL UNIVARIATE ANALYSIS

Analyzing 'loan_amnt’



Most of the customers applied for a loan amount between 5000 and 10000

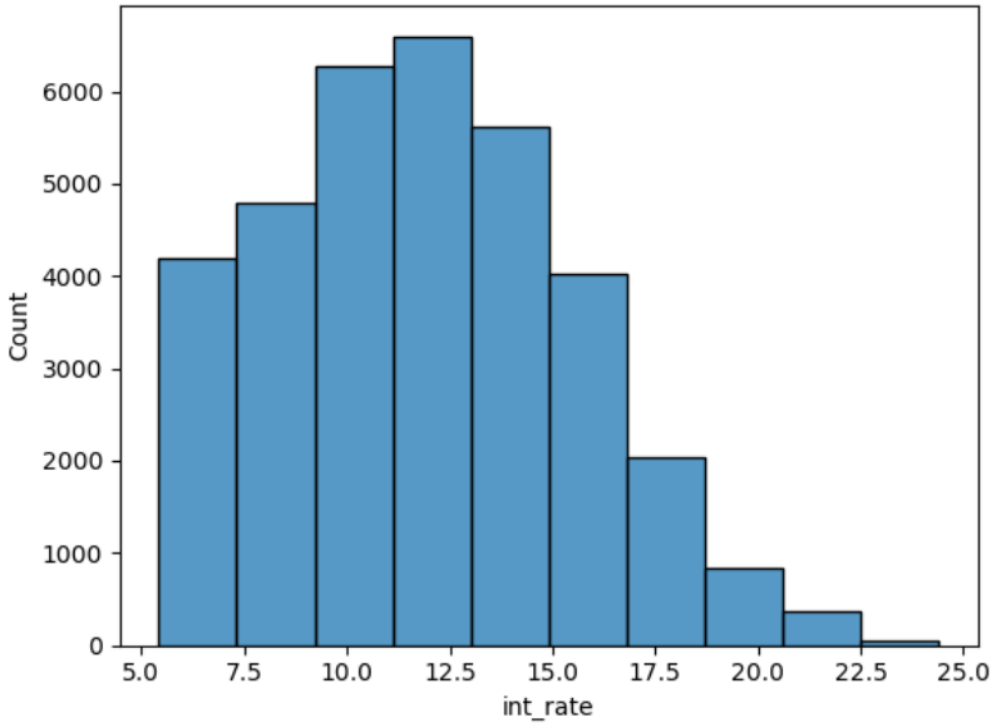
Analyzing 'term'



This shows that most of the customers apply loan for a term of 36 months

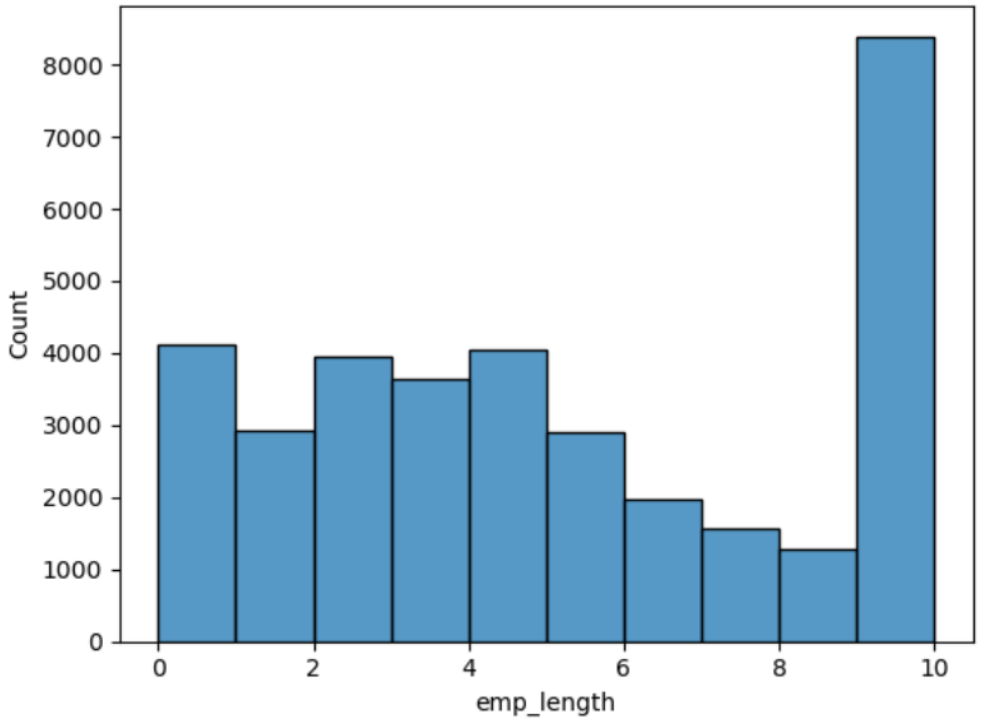
NUMERICAL UNIVARIATE ANALYSIS

Analyzing 'int_rate’



There are more loans from 10% to 15% interest rates

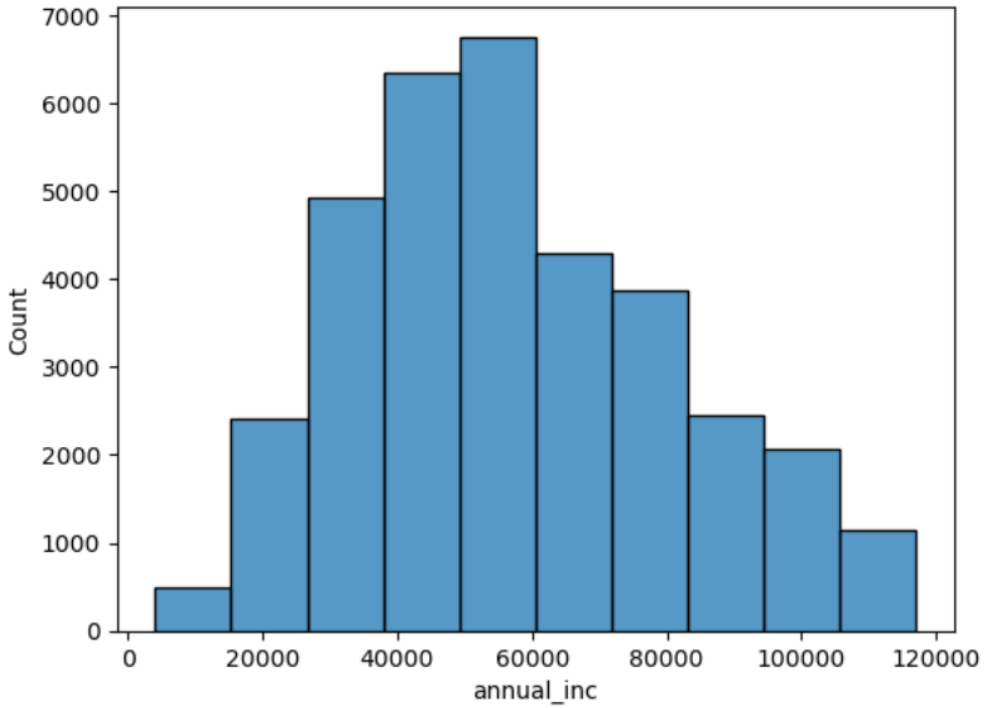
Analyzing 'emp_length'



Most of the customers have an experience of around 10 years

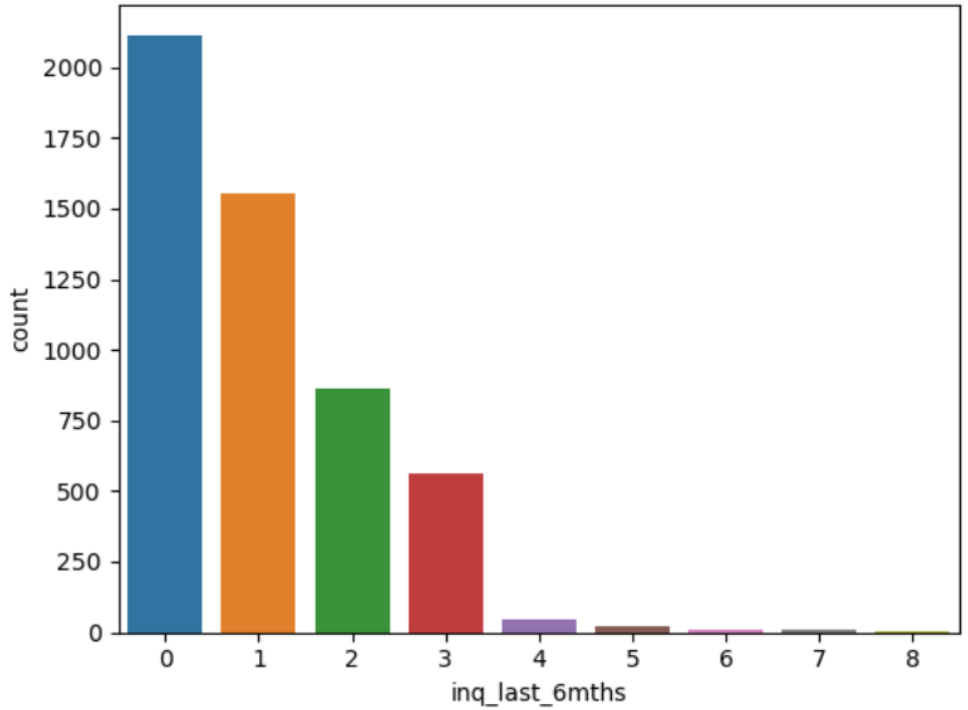
NUMERICAL UNIVARIATE ANALYSIS

Analyzing 'annual_inc’



Most of the customers have an annual income between 35000 to 60000 (approx.)

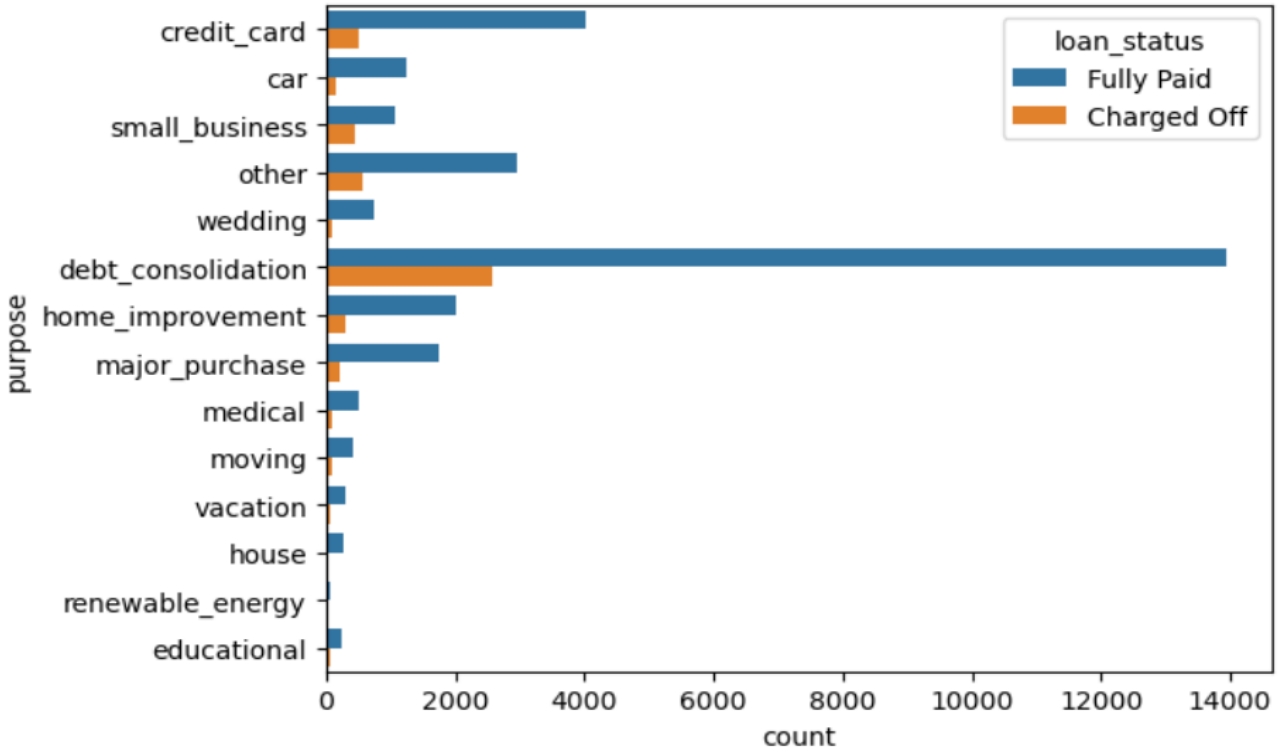
Analyzing 'inq_last_6mths'



As the number of enquires in last 6 months are less, the count of defaulters are more

SEGMENTED UNIVARIATE ANALYSIS

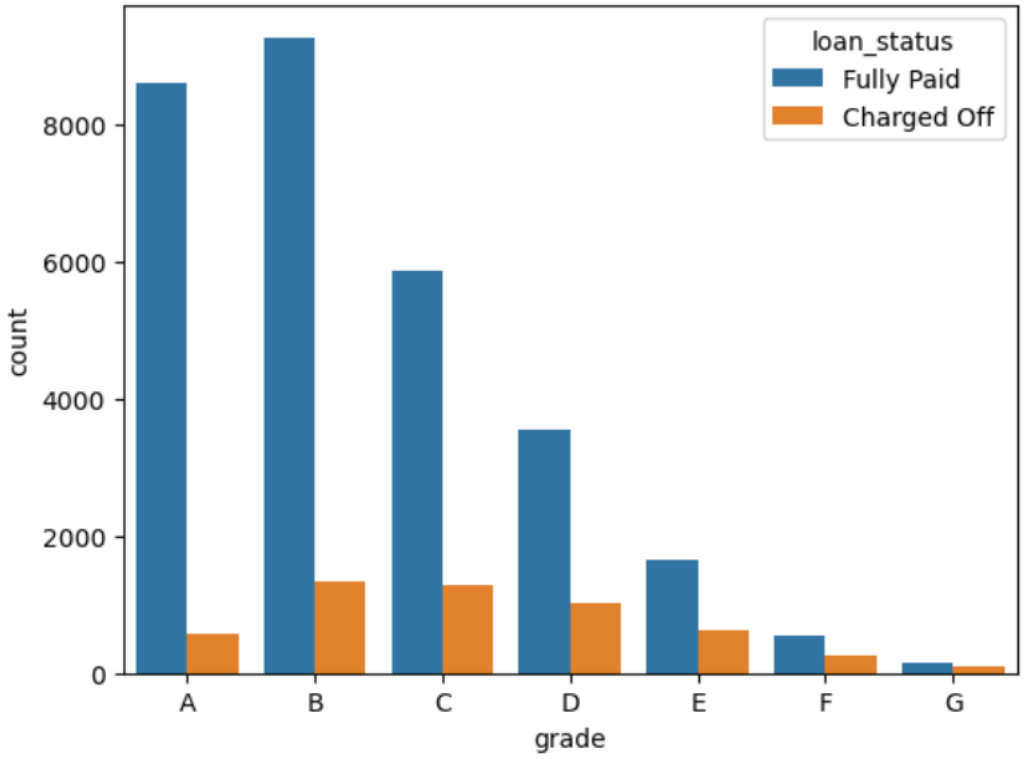
‘loan_status’ based on ‘purpose’



There are more loans in the category 'debt_consolation'. So, there can be more amount of fully paid and charged off customers

SEGMENTED UNIVARIATE ANALYSIS

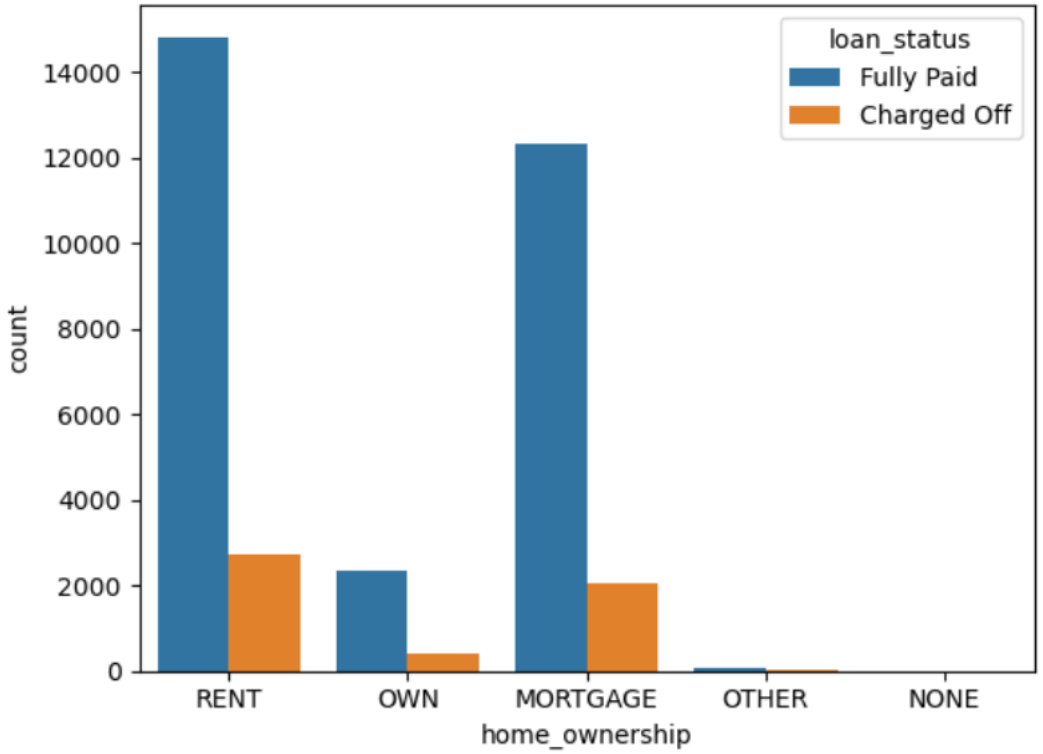
'loan_status' based on 'grade'



The count of fully paid and charged off customers decreases as grade increases. Most of the Grade A customers pay the loan fully and they are less likely to default. Meanwhile, B has more customers. Despite having the maximum count of fully paid customers, they have more defaulters.

SEGMENTED UNIVARIATE ANALYSIS

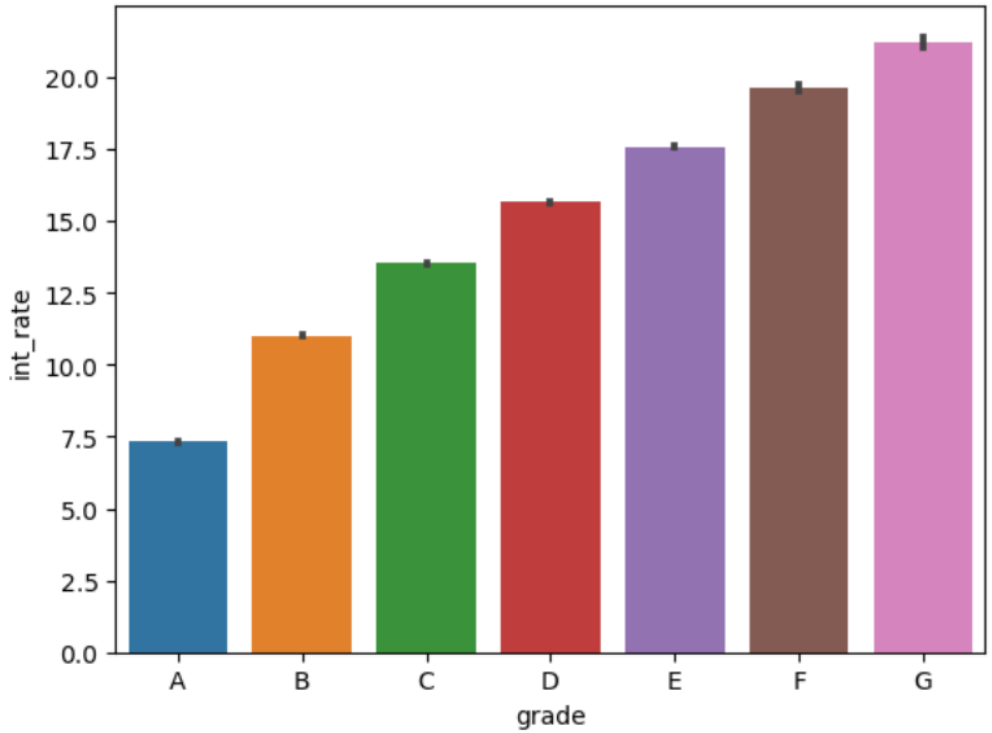
‘loan_status’ based on ‘home_ownership’



The customers who have a rented and mortgaged house are more likely to default even if the count of fully paid is high when compared to own house customers

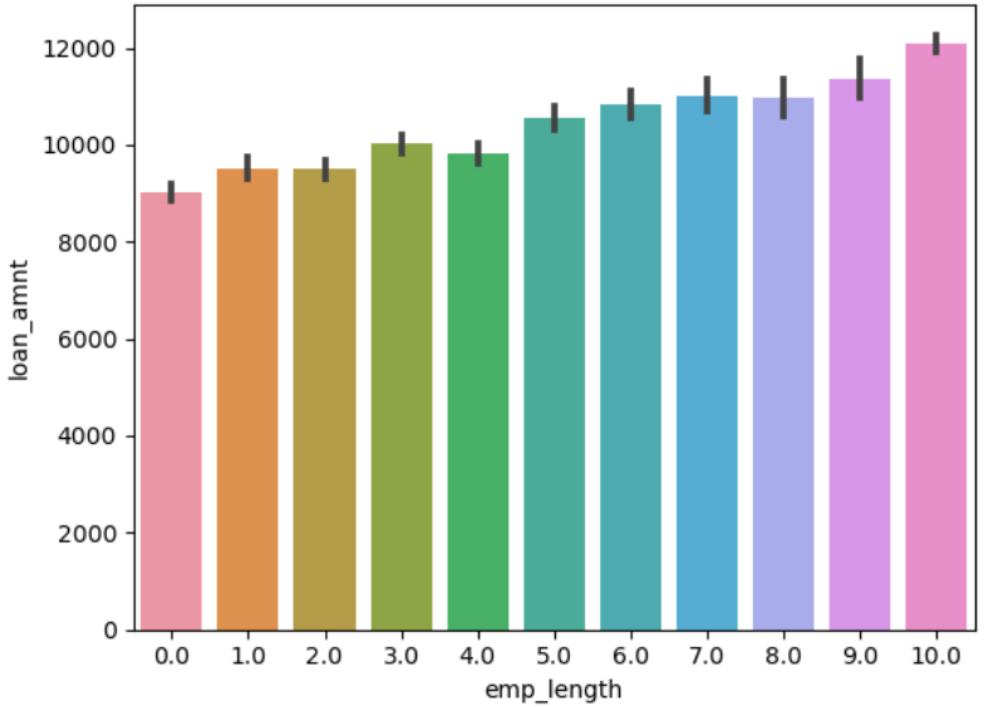
BIVARIATE ANALYSIS

‘grade’ vs ‘interest_rate’



We can observe that as grade increases, the interest rate increases

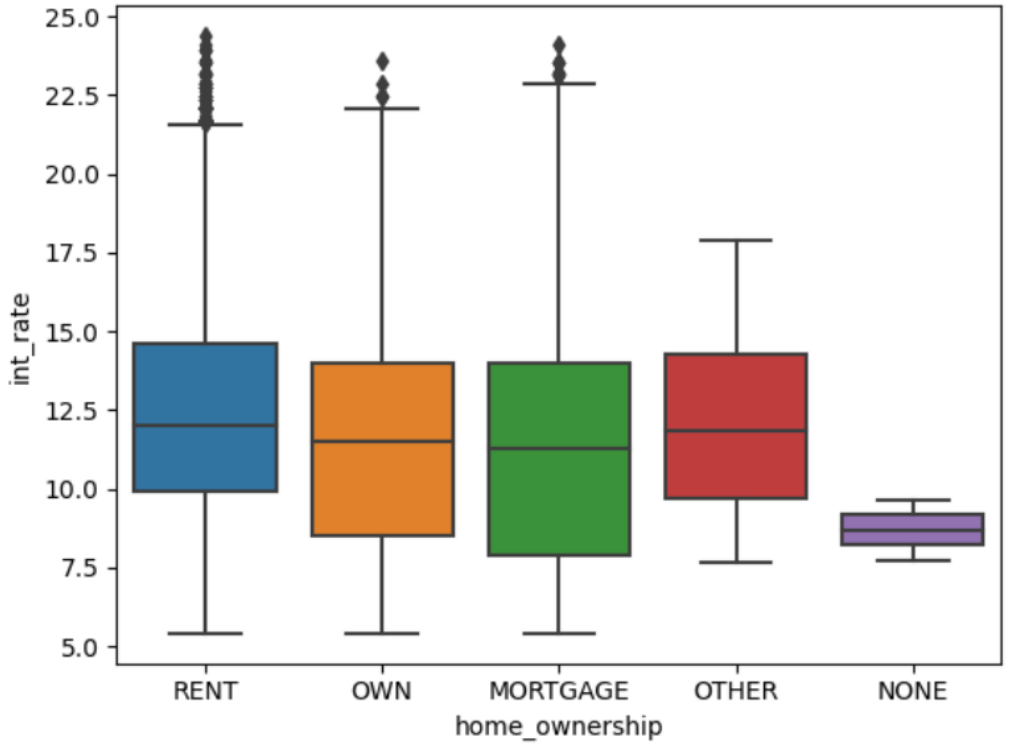
‘loan_amnt’ vs ‘emp_length’



It is observed that as experience increases, the loan_amnt also increases. Aged customers are likely to get loans of higher amount

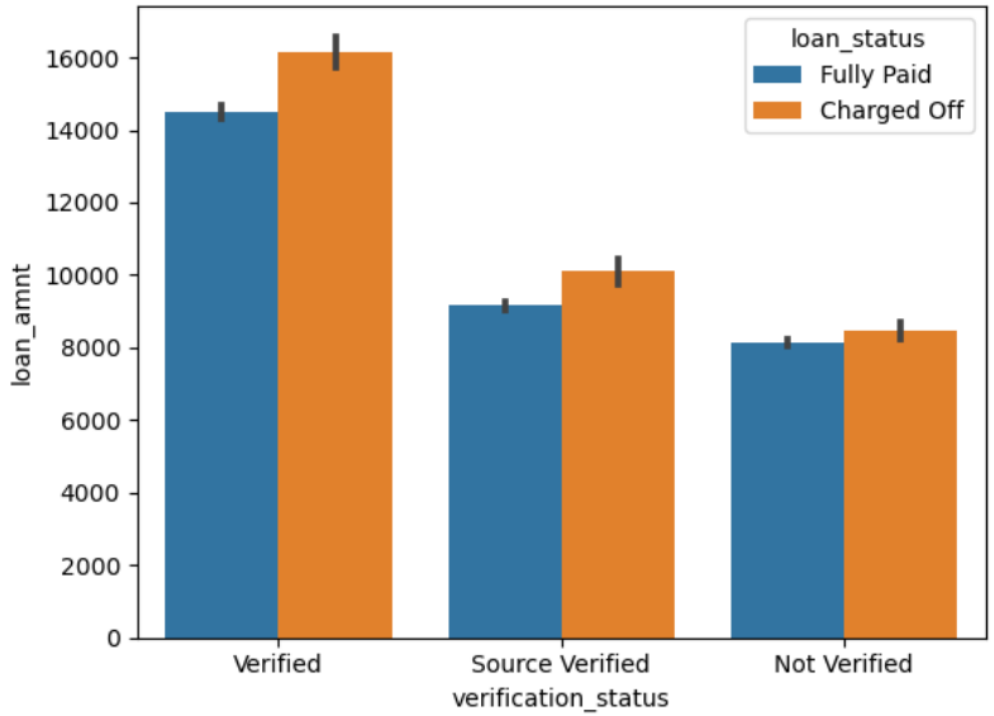
BIVARIATE ANALYSIS

‘home_ownership’ vs ‘int_rate’



Among all, the customers who have mortgaged house pay more interest and they have more chances to become a defaulter

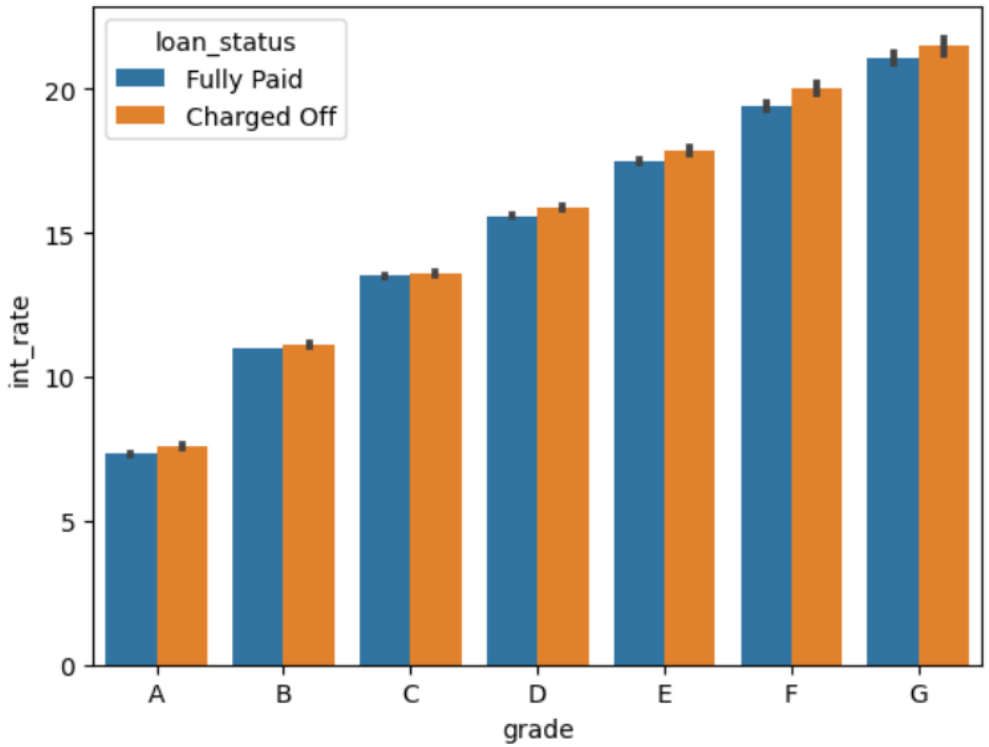
verification_status vs loan_amnt



It can be said that a customer is likely to be a defaulter if he/she is verified and the loan amount is more than 14000

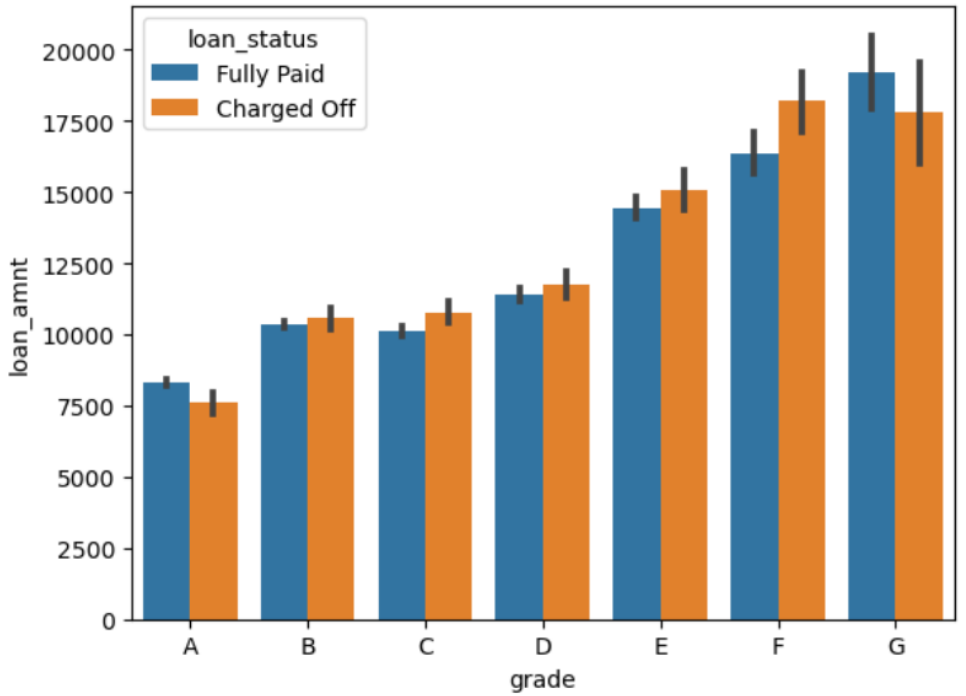
BIVARIATE ANALYSIS

‘grade’ vs ‘int_rate’



It was already seen that as grade increases, interest rate increases. Looking at the above plot, it can be inferred that defaulters are more likely to be in case of Grade G customers who pay interest > 20%

‘grade’ vs ‘loan_amnt’



Grade F customers who pay loan amount between 15000 to 20000 are likely to be defaulters

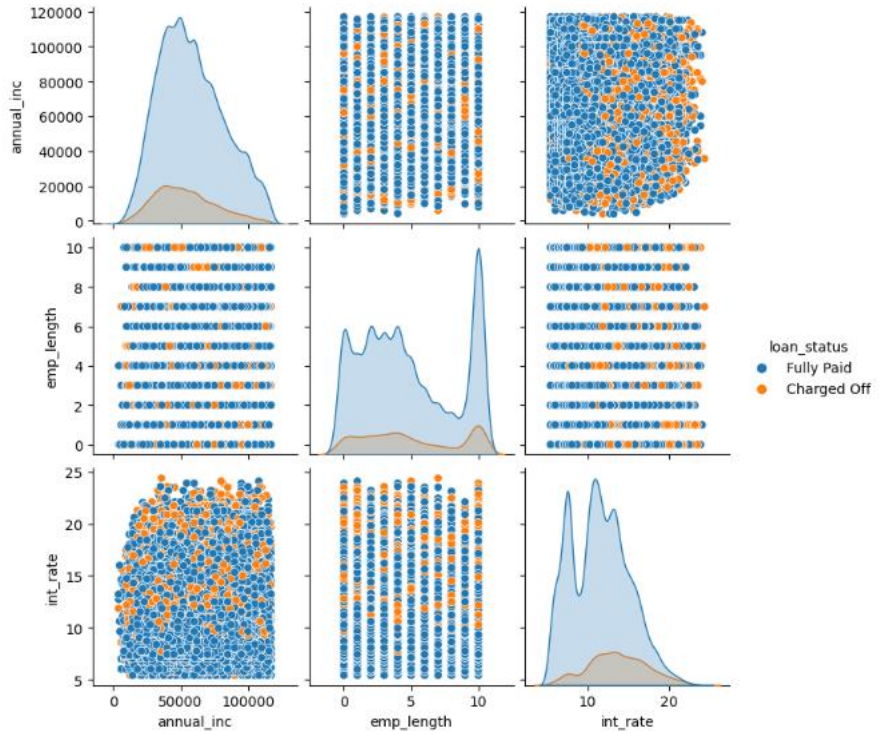
MULTIVARIATE ANALYSIS

- Heat map- to check the correlation

1) As annual income increases, the loan amount increases¶
2) As the term increases, the interest rate increases
3) As 'revol_util' increases, the interest rate increases
4) As interest rate increases, 'pub_rec_bankruptcies' increases
5) As the 'emp_length' increases, annual income increases

MULTIVARIATE ANALYSIS

- Analyzing with a pairplot



As the emp_length increases, the annual income of the employee increases. So, they will be in a position to pay the loan easily

INFERENCE

- Based on grade:
  - There is a higher chance for defaulting in case of
    - Grade G and interest rate > 20%
    - Grade F and amount between 15k and 20k
- As the interest rate increases, the chance of defaulting also increases
- Customers with more emp_length (experience) would have a higher annual income and are more likely to pay the loan
- Customers with public records and bankruptcies are more likely to default
- Customers having an own house are less likely to be defaulters. (More chance- Rented and Mortgaged customers)
- Verified customers with a loan amount more than Rs. 14000 have a higher chance of defaulting
- Customers getting a loan for ‘debt consolidation’ are more likely to default
- Probability of defaulting is high towards the end of the year (Maximum in December)