

Lending Club Case Study

Sritharini Radhakrishnan

&

Sirigireddy Siva Teja Reddy

Road map

1. **Data Sourcing**
2. **Data cleaning**
3. **Univariate Analysis**
4. **Segmented univariate analysis**
5. **Bi-variate analysis**
6. **Multi-variate analysis**
7. **Derived metrics**
8. **Inference & Recommendation**



1. Data Sourcing

Data sets

1. **Loan.csv** - The main dataset
2. **Data_Dictionary.csv** - Definition of all the columns that are present in the loan dataset.



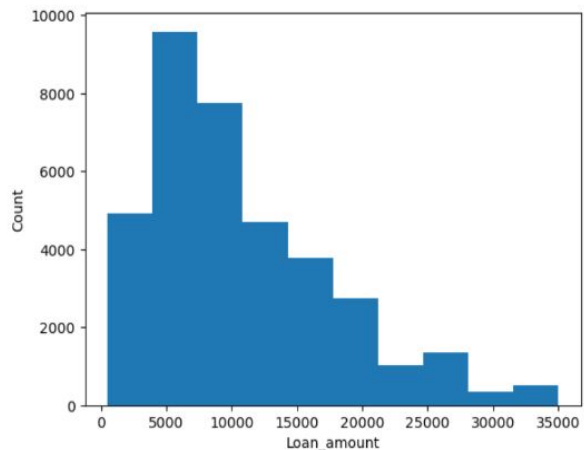
Data cleaning

- A. Cleaning data(null valued columns, one valued columns, unique identifier columns)
- B. Standardising values(to int, to float, to percentage)
- C. Dropping duplicate values
- D. Segregating numerical and categorical data
- E. Missing value imputation(mean/median/mode)
- F. Filtering data
- G. Outlier treatment



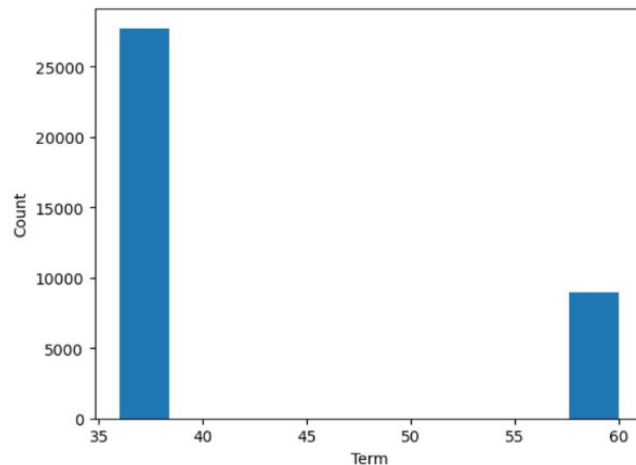
3. Univariate Analysis

Loan



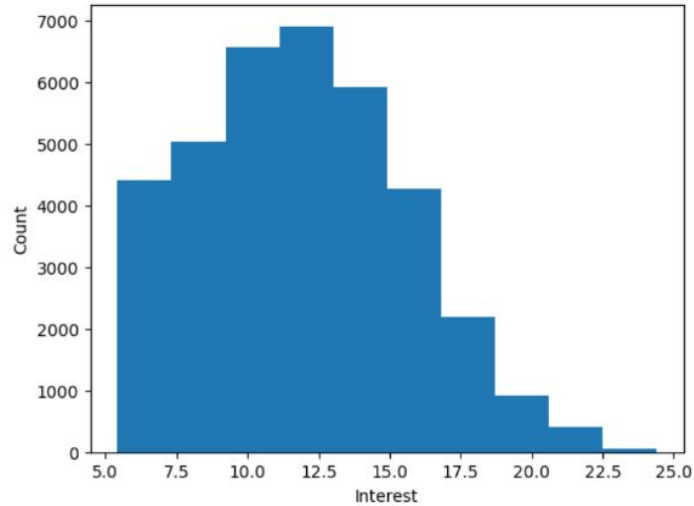
The loan amount varies from 500 to 35000 with the mean of 10682
The maximum number of people took loan for amount 5000 to 10000

Term



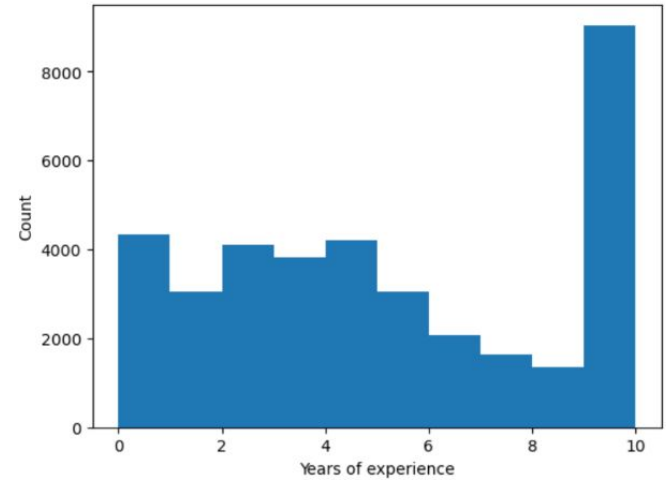
From the histogram we can clearly see that the customers
prefers 36 months over 60 months
Thich means that they would like to payback loan as soon as possible.

Interest



The interest rate ranges from 5.42 to 24 %
10 to 15 percent looks like the most commonly lend interest rate

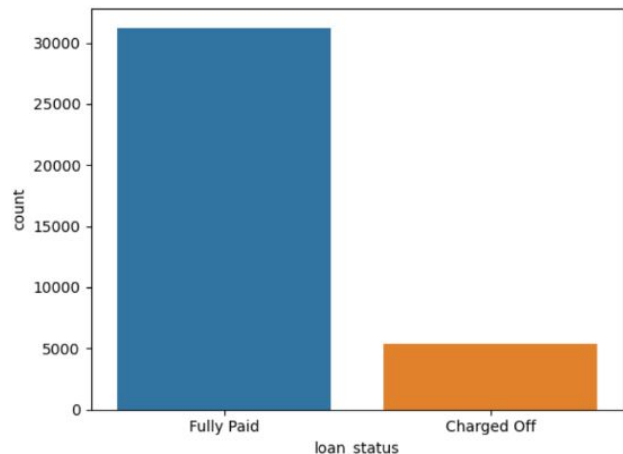
Experience



The number of years of experience ranges from 0 to 10 and
the maximum number of customers have 10 years of experience

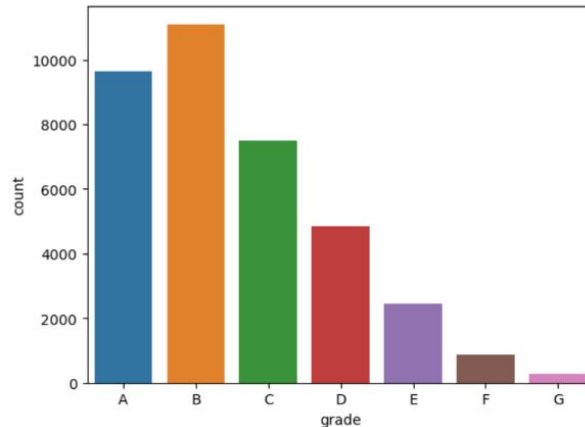
Categorical univariate analysis

Loan_status



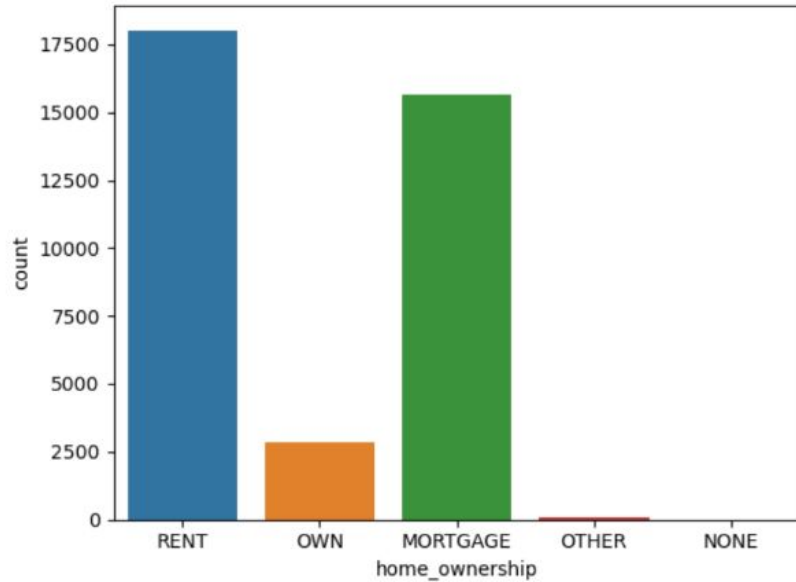
Most of the customers have paid back the loan completely ,
14.78% of the customers are defaulters

Grade



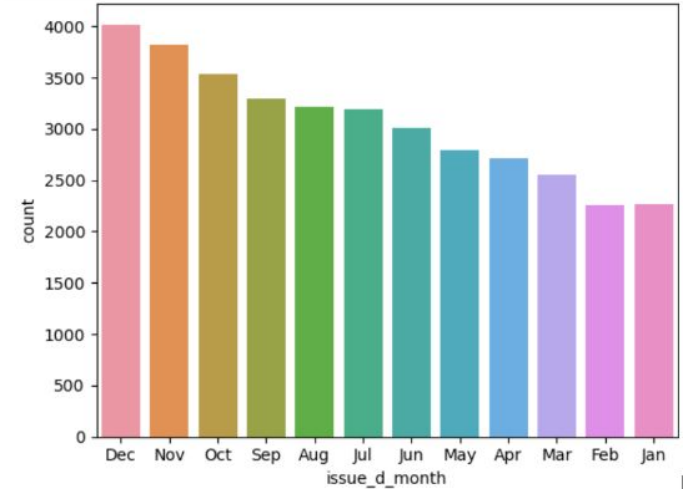
We can see that as the grade increases through A to G,
the number of customers decreases
Most of the customers belongs to A,B,C grade

Home_ownership



Most of the customers lives in rented house or paying mortgage

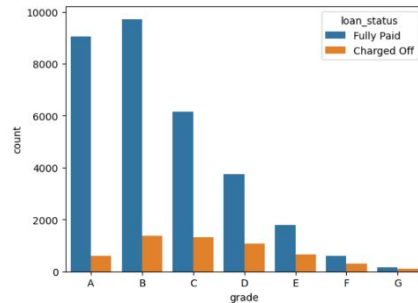
Loan issued month



The more number of loans are issued in the last quarter compared to any other quarter

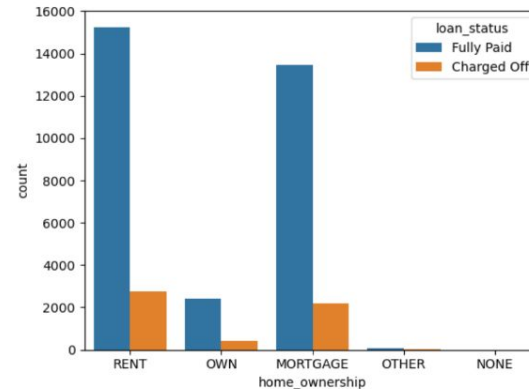
4. Segmented univariate analysis

Loan status based
On grade



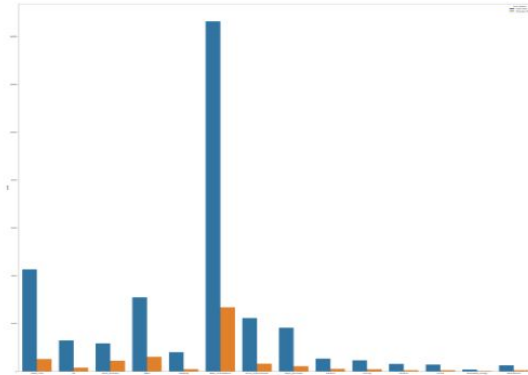
- We can see that as the grade increases, the loans fully paid and charged off decreases
- The good candidate for loan is A grade customers since their payback is more and they are less likely to default
- The number of employees who took loan in B grade are more, which also gives higher chances of full repayment and defaulting as well.

Loan status based
On home_ownership



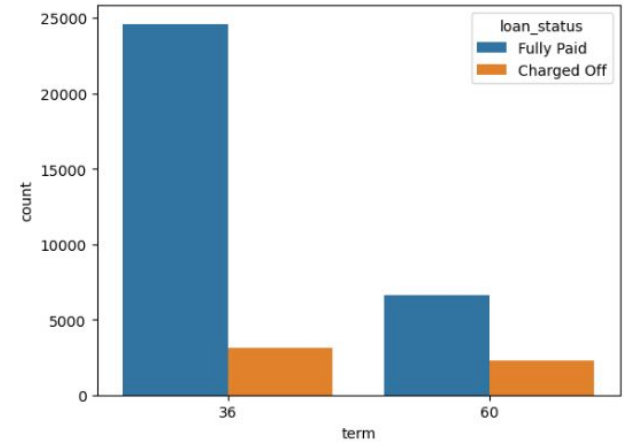
- The customers who own house are less likely to default compared to rented house customers and mortgage payers
- We can also hold own house as a collateral which encourages customers to pay back completely

Purpose based on loan



- The most common purpose of loan is debt_consolidation
- Therefore chances of fully paid and defaulting is also more in this criteria

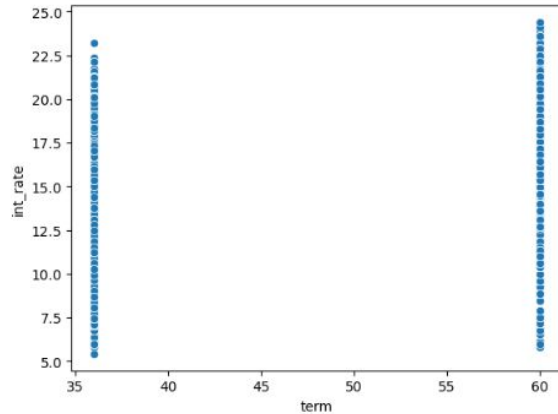
Term based on loan status



The ratio of fully paid to charged off is much lesser for customers who took 36 months emi

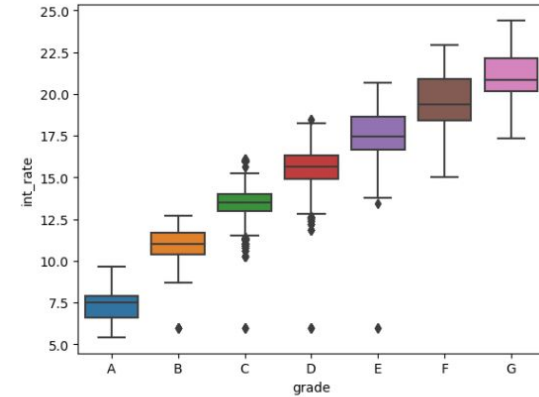
5. Bivariate Analysis

Term vs Emp_length



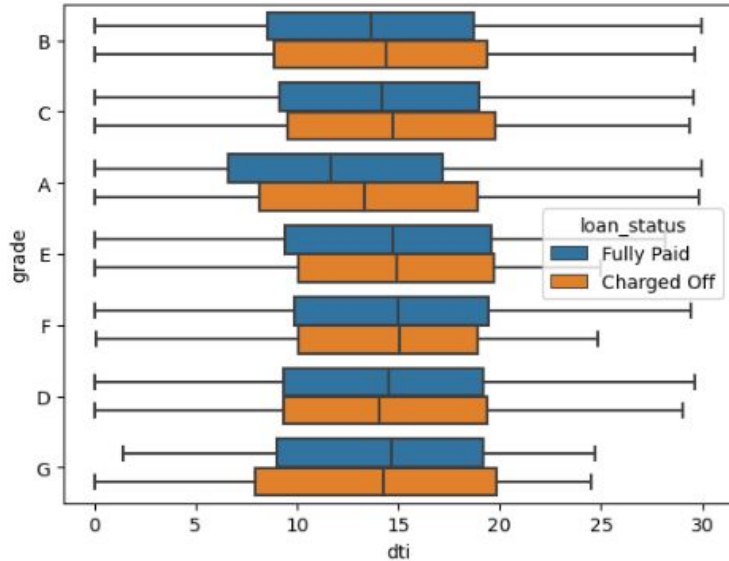
we can observe that the interest rate is independent of the number of months the customer chooses to payback the loan

Int_rate vs grade



As the grade increases, interest rate increases

Grade vs dti



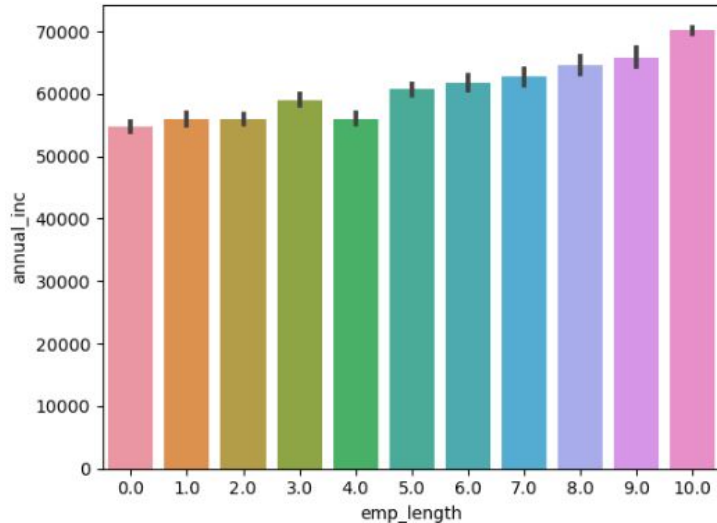
dti is more for lower grade employees, and that results in defaulting

Grade vs annual_income

annual_inc	
grade	
A	55800.0
B	55000.0
C	54263.0
D	56000.0
E	60000.0
F	67200.0
G	72800.0

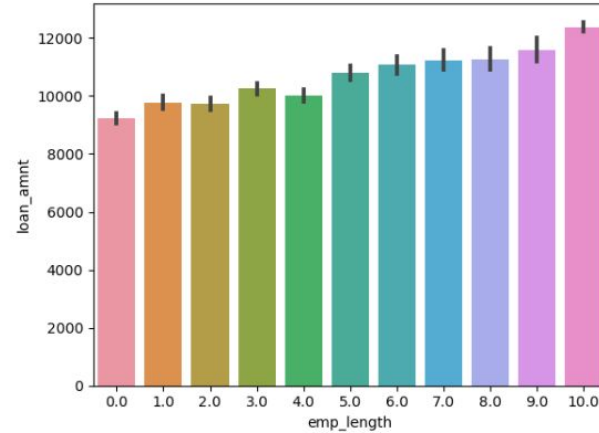
As the grade increases, salary increases, which makes it safe to assume, G is the highest grade

annual_inc vs emp_length



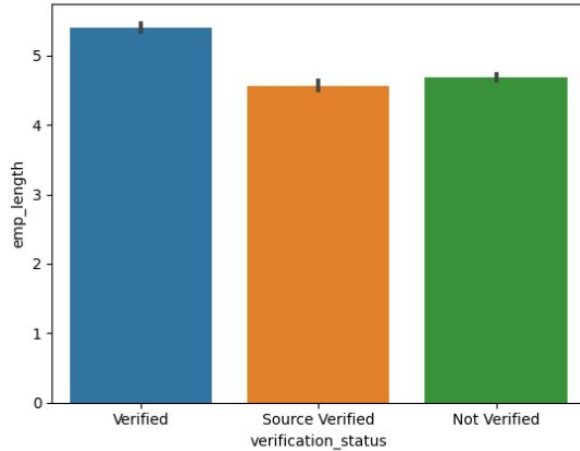
As the experience of an employee grows,
their salary grows as well

Loan_amnt vs emp_length



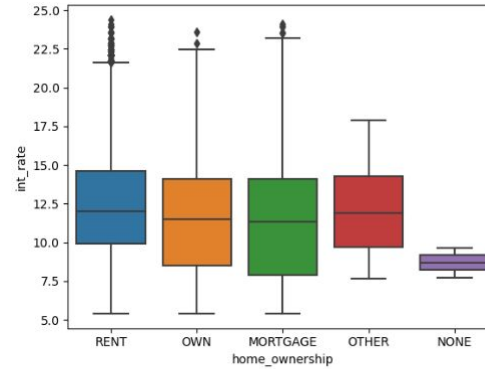
- As the employee length increases, the loan amount increases
- Increase in experience indicates increase in age which leads to increase in need for more valued possessions
- This leads to increase in the amount of the loan taken

Emp_length vs verification_status



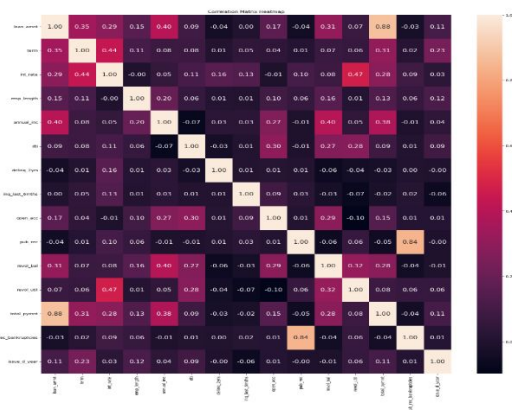
The customers who have their incomes verified have more experience

Int_rate vs home_ownership

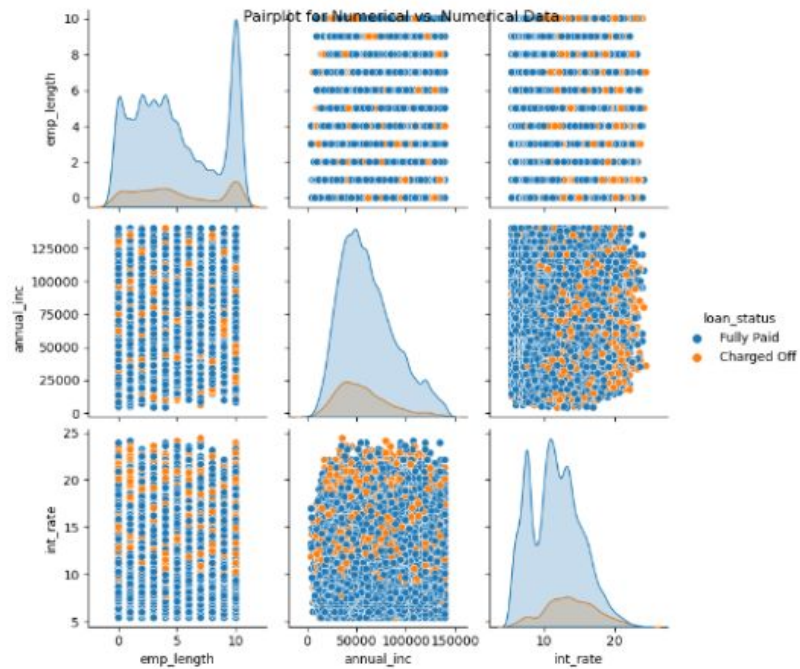


- customers who pays mortgage pays higher interest compared to customers who owns house
- customers who pays mortgage most likely to default

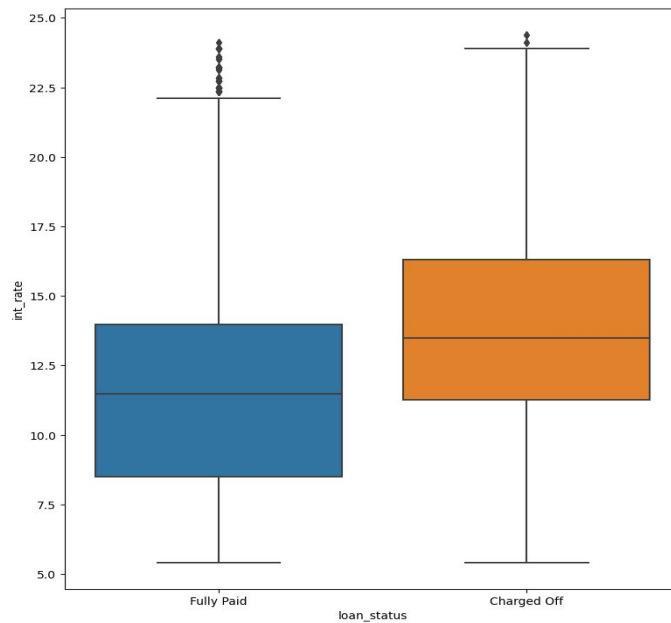
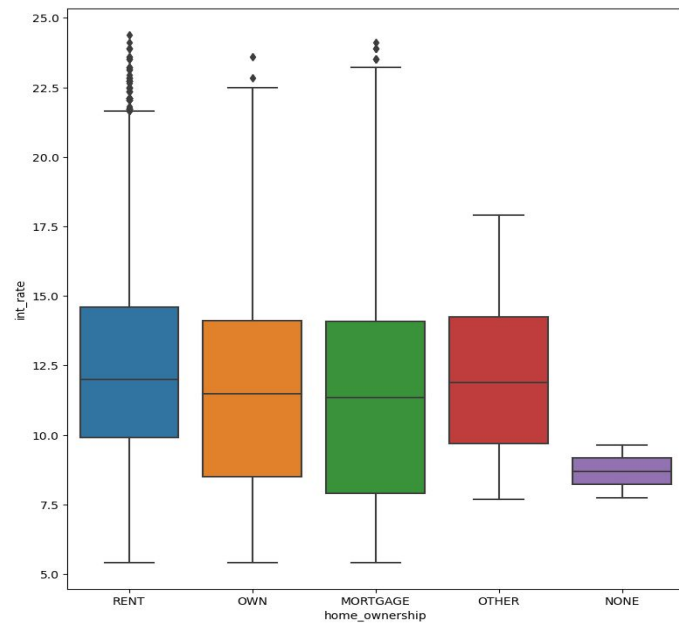
6 . Multivariate Analysis



- revol_util and int_rate are positively correlated, as revol_util increases, int_rate increases
- pub_rec and int_rate are positively correlated, as pub_rec_bankruptcies increases, pub_rec increases,
- This indicates that people who has bankruptcies, tend to default
- annual_income is positively correlated to loan_amount
- more the income, more chance for increased loan lending
- annual income increases, revolving balance increases



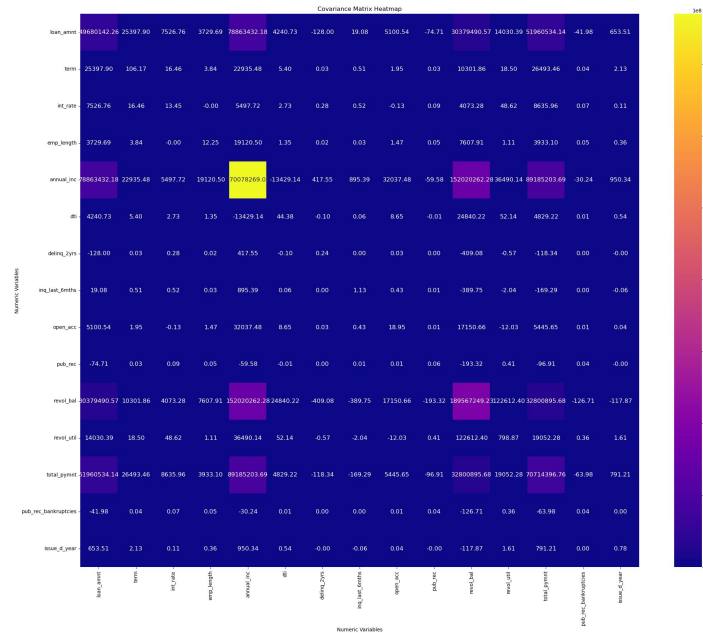
When experience of employee increases, salary of the customer increases, which increases the chances of full payback



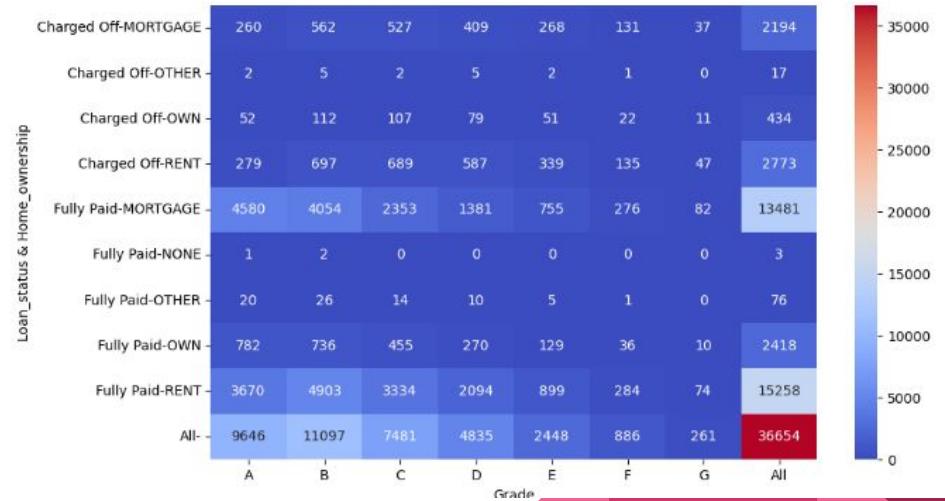
7. Derived metrics

Most of the customers who pays mortgage are of B,C,D grade employees.

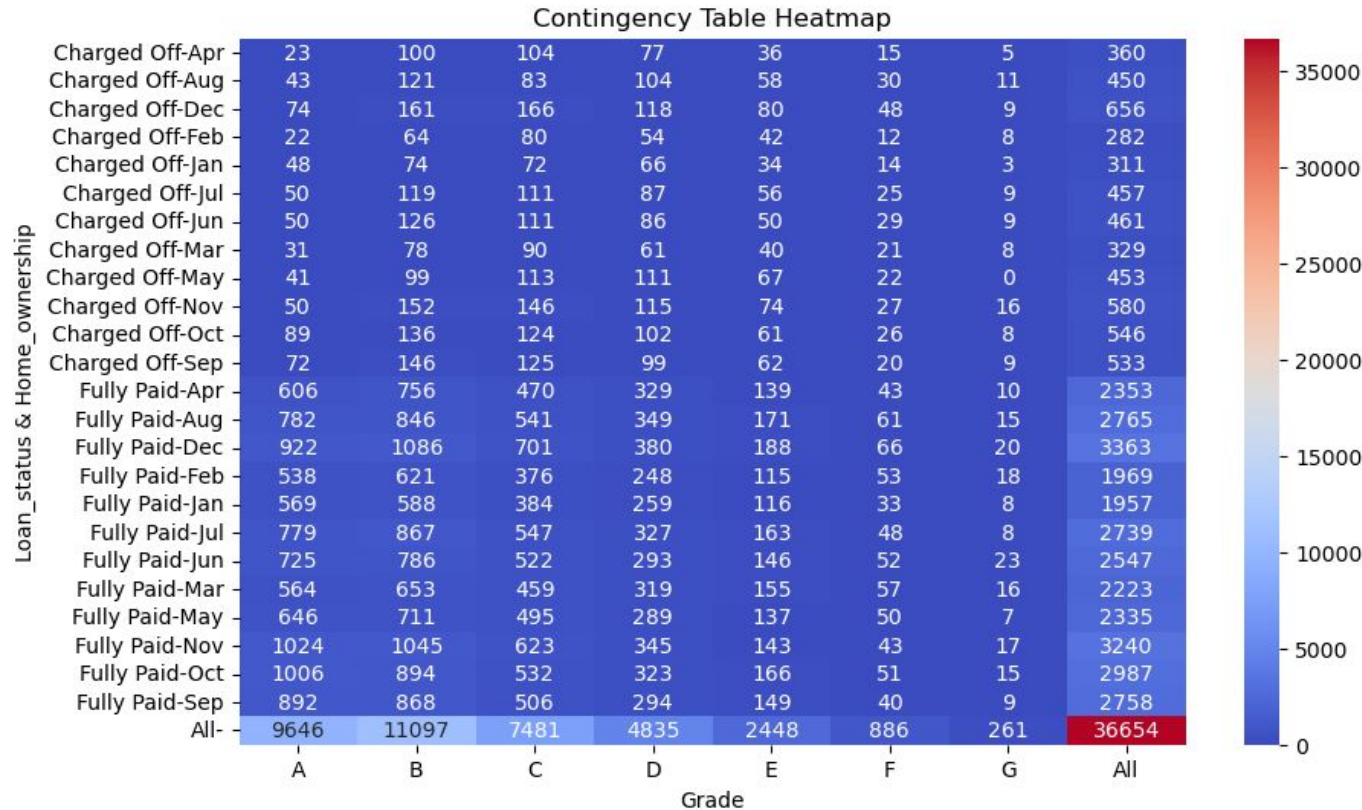
Covariance matrix



Contingency heat map



Based on the contingency heat map, the last quarter has more number of charged off customers



The major driving factors for defaulters and to prevent financial loss:

1. int_rate
2. annual_income
3. emp_length
4. pub_rec_bankruptcies
5. dti
6. home_ownership
7. grade



8. Inference & Recommendations :-

1. B grade employee takes more loans, results in more defaulters, so B grade is more likely to be defaulters
2. More the interest rate, its more likely to cause defaulting.
3. Customers who pays mortgage pays higher interest, so mortgage payers are more likely to be defaulters
4. When experience of employee increases, salary of the customer increases, which increases the chances of full payback
5. Customers with public records and bankruptcies are most likely to be defaulters
6. The number of loans issued is more in the last quarter, compared to any other quarter, this might be because the lending club might have deadline to achieve at the year end, so the last quarter is more prone to defaulters
7. Similarly, our contingency heatmap in derived metrics says that there more number of defaulters in the last quarter,
Therefore, the loans lended in the last quarter has the highest possibility of defaulters.
8. dti is more for A,B,C grade employees, which leads to more interest, which more likely cause defaulting





Thankyou