Lending Club Case Study

Gaurav Joshi

Jitesh Khurana

Contents

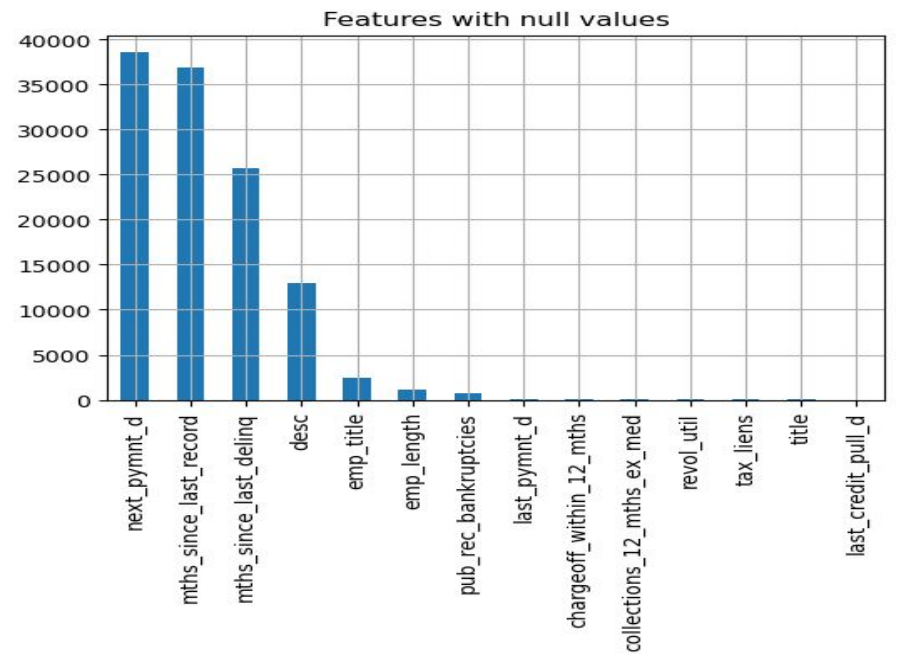
Problem Statement

Analyze and Identify trends and patterns to capture risky loan applicants who are likely to default in order to minimize the risk of losing money while lending money to customers by consumer finance company.

Balance between loan application rejection to optimize monetary loss and business loss
-

- 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

Data Cleanup



Achieved in Cleanup -

1. Removed columns where all values were null

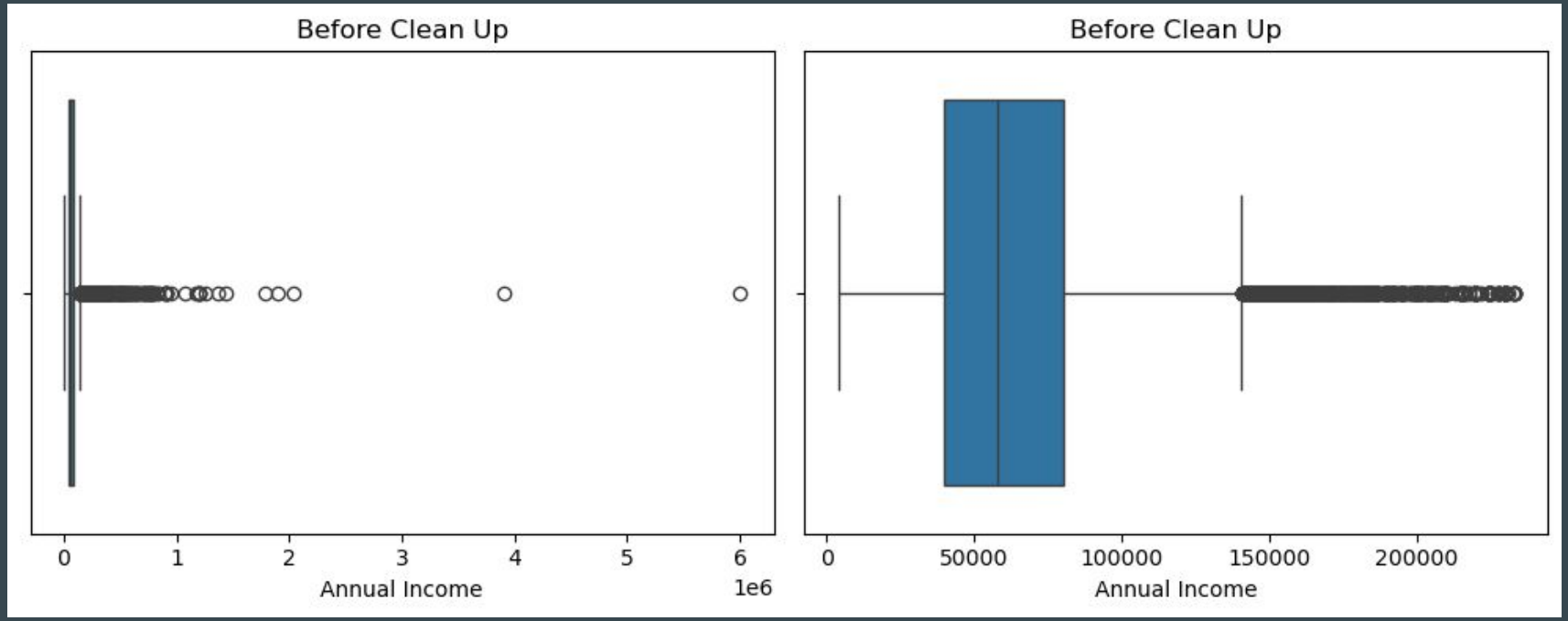
2. Removed columns where all rows had duplicate values

3. Filled null values for emp_length - by mode

4. Removed columns which were only representing unique IDs

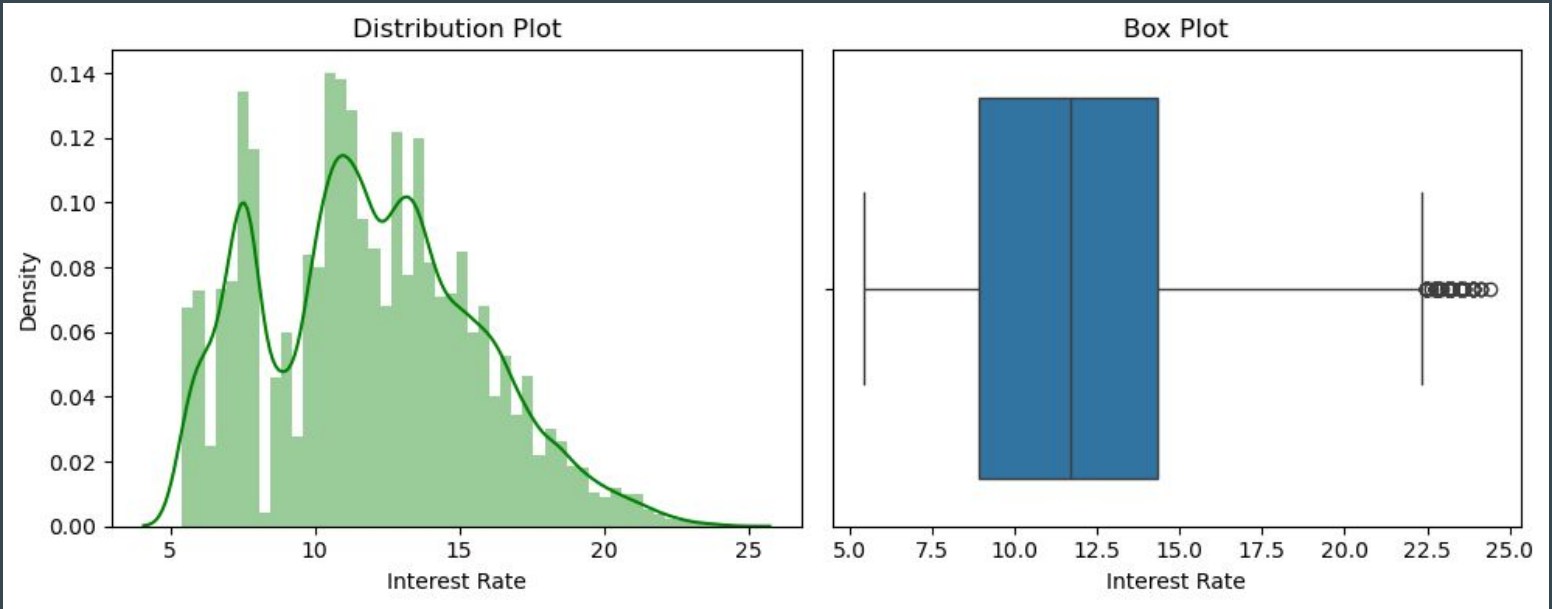
5. Removed columns which ideally are part of post loan approval

Univariate Analysis - Annual Income



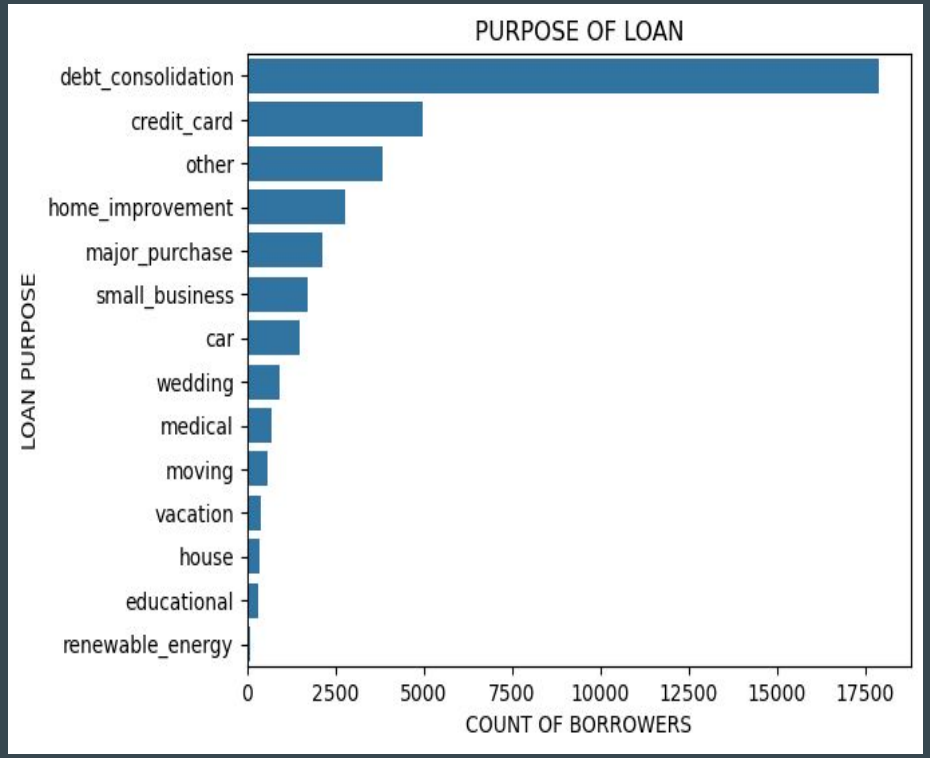
- Removed outliers from Annual Income - picked 99th percentile as threshold since it was adding significant bias to the data

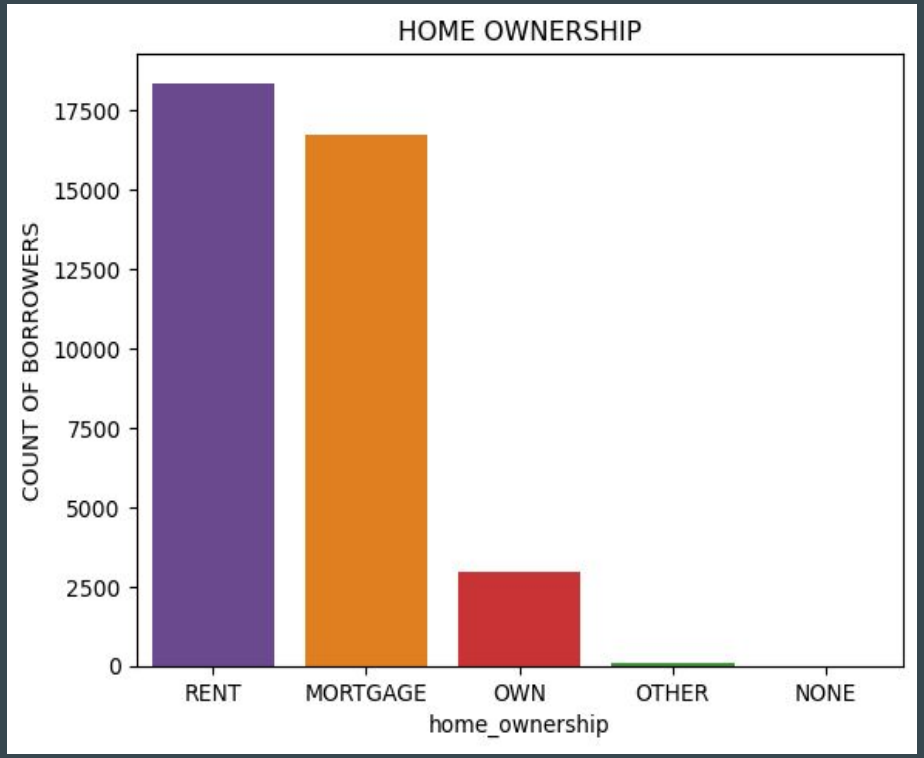
Univariate Analysis - Annual Income



- The average interest rate is 12%, post 75 percentile the interest rate increases greatly
- Most of the borrowers prefer to get loan at interest rate ranging between 9% to 14%

Univariate Analysis - Categorical Variables

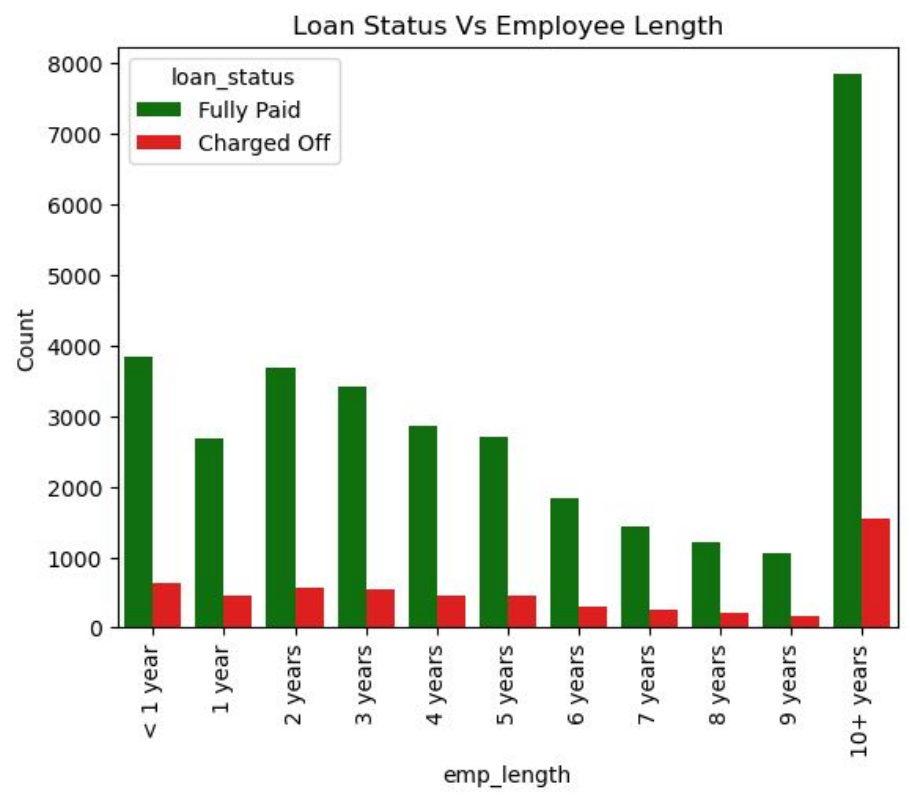




Univariate Analysis - Summary

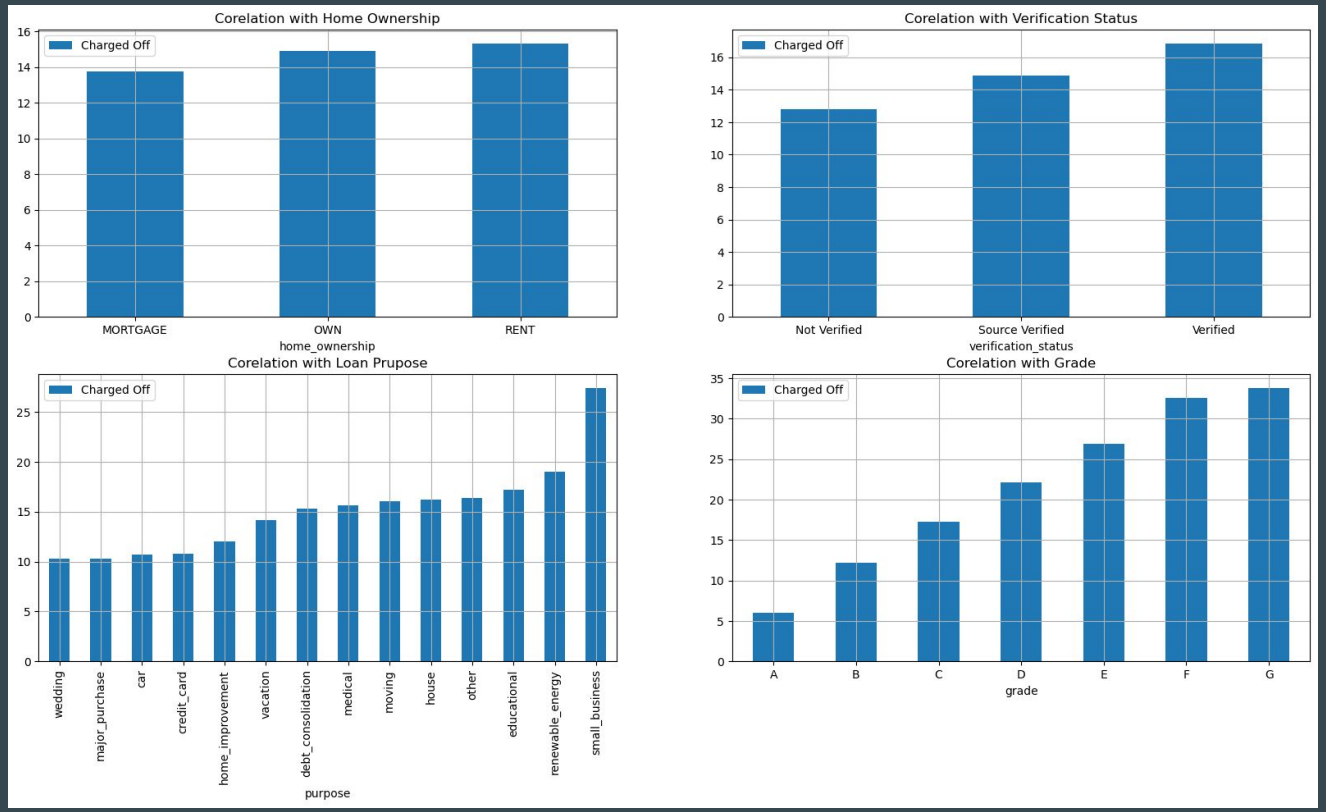
- Annual Income had outliers in the upper fence which were removed.
- There was a huge jump between 99 and 100 % of data in Annual Income and this was removed.
- The average interest rate is 12% but there is huge jump after 75 percentile.
- Most of the borrowers get interest rate between 9% to 14%.
- Most of the loans are taken for debt consolidation and credit card bill payment.
- Very less percentage of borrowers have their own house
- Outlier presence was not impacting the fund amount and loan amount

Bivariate Analysis - Loan Status against employment length

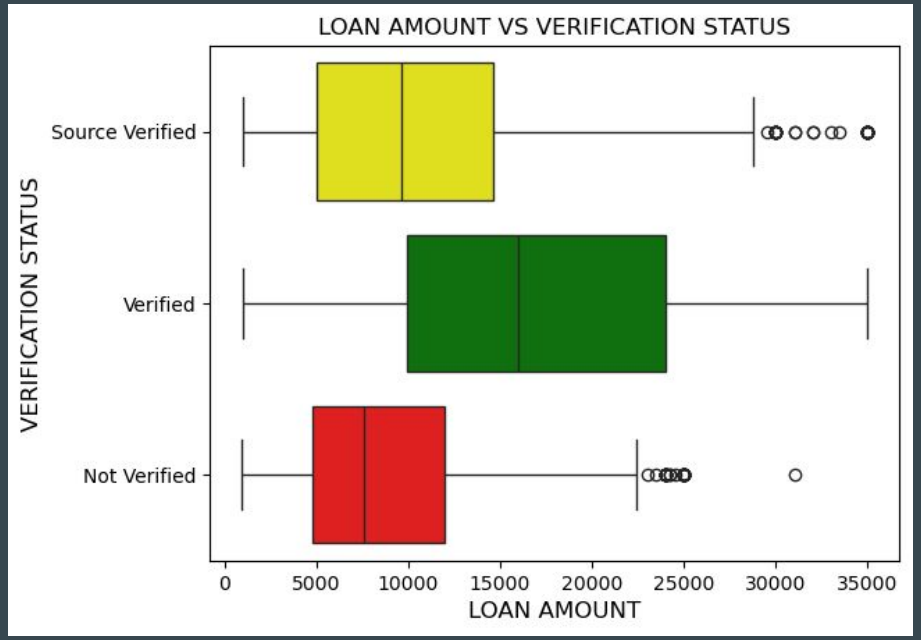


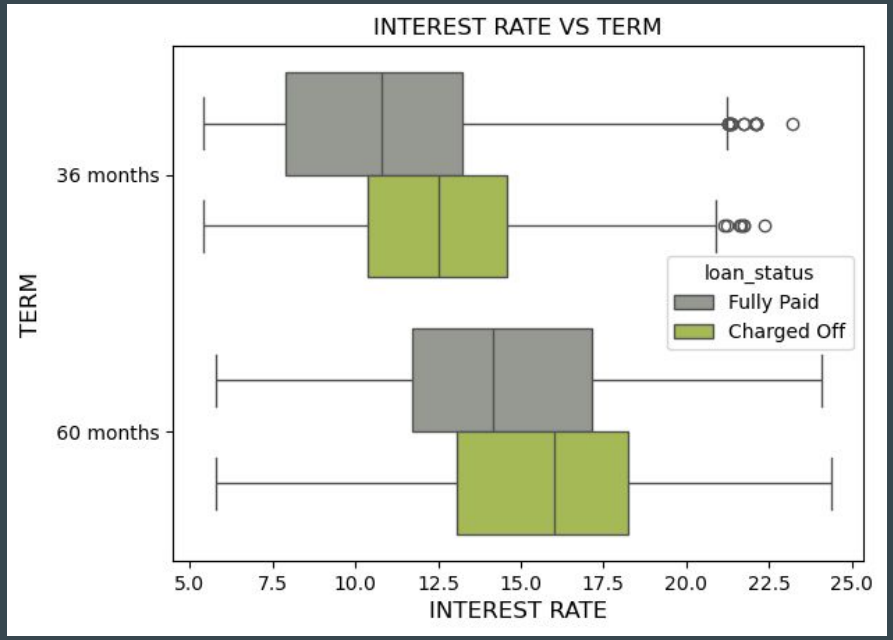
loan_status	Charged Off	Fully Paid	Ratio
emp_length			
10+ years	1541	7835	19.67
7 years	262	1432	18.30
1 year	452	2685	16.83
5 years	455	2709	16.80
8 years	202	1215	16.63
< 1 year	631	3836	16.45
6 years	303	1845	16.42
3 years	551	3421	16.11
4 years	454	2855	15.90
2 years	560	3679	15.22
9 years	157	1058	14.84

Bivariate Analysis - Loan Status against other critical features

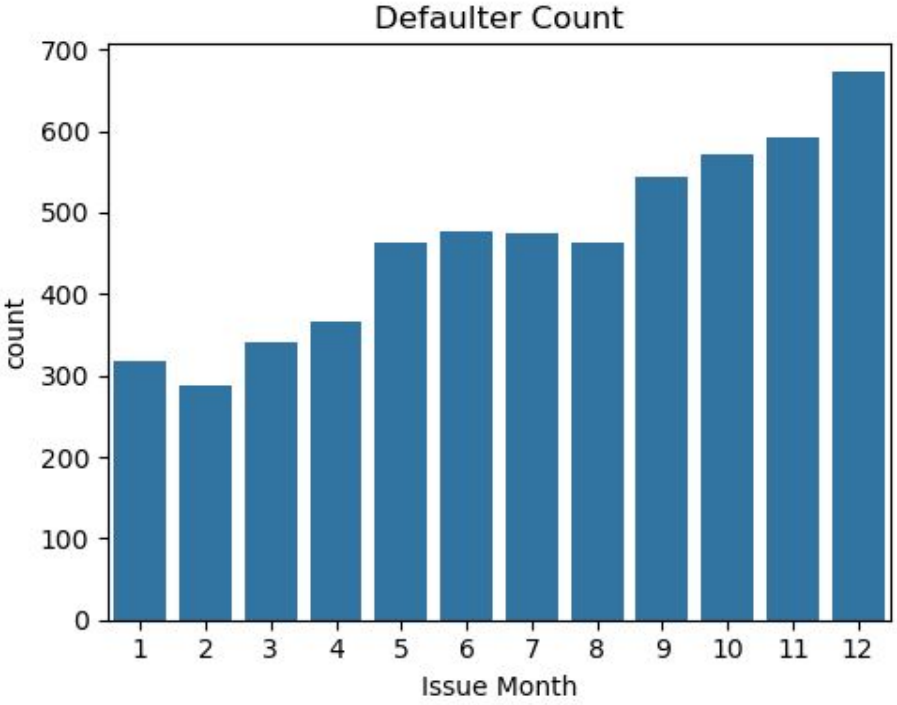


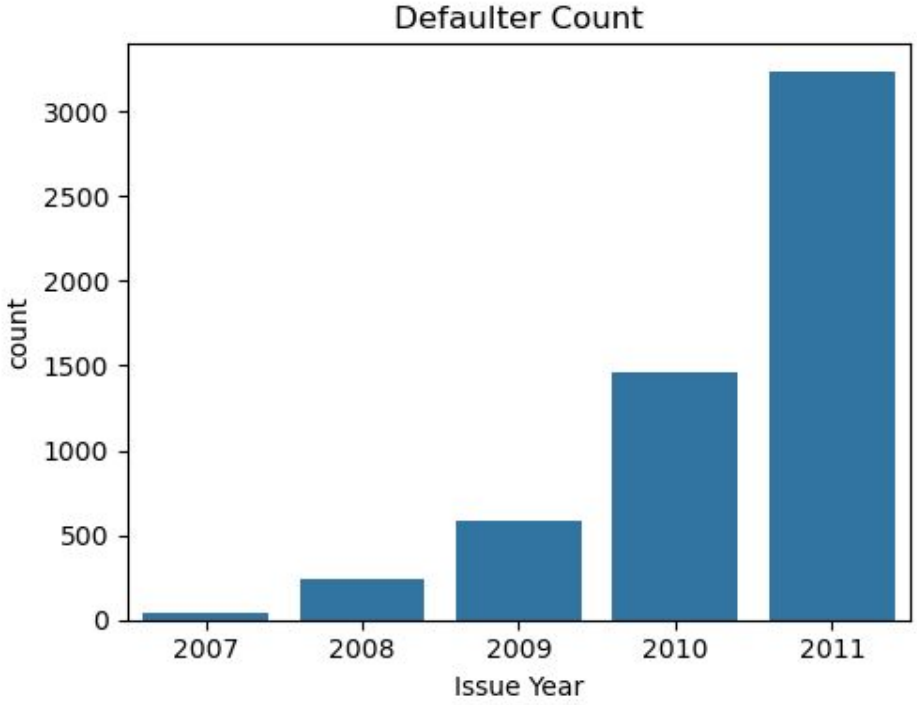
Bivariate Analysis - Loan Amount * Verification and Interest * Term



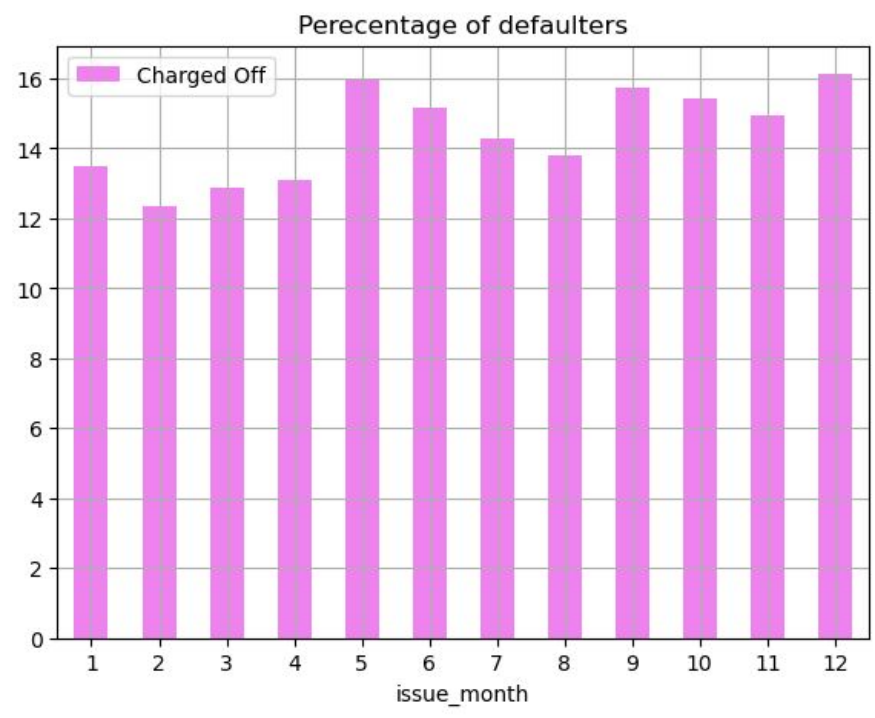


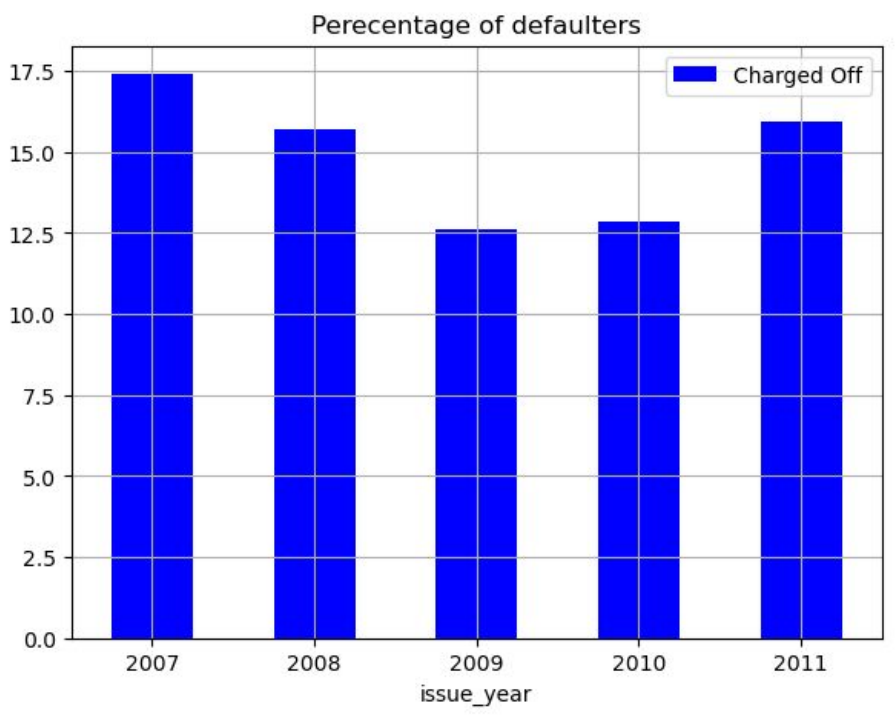
Bivariate Analysis - Charged Off vs month-year





Bivariate Analysis - Charged Off Percentage vs month-year

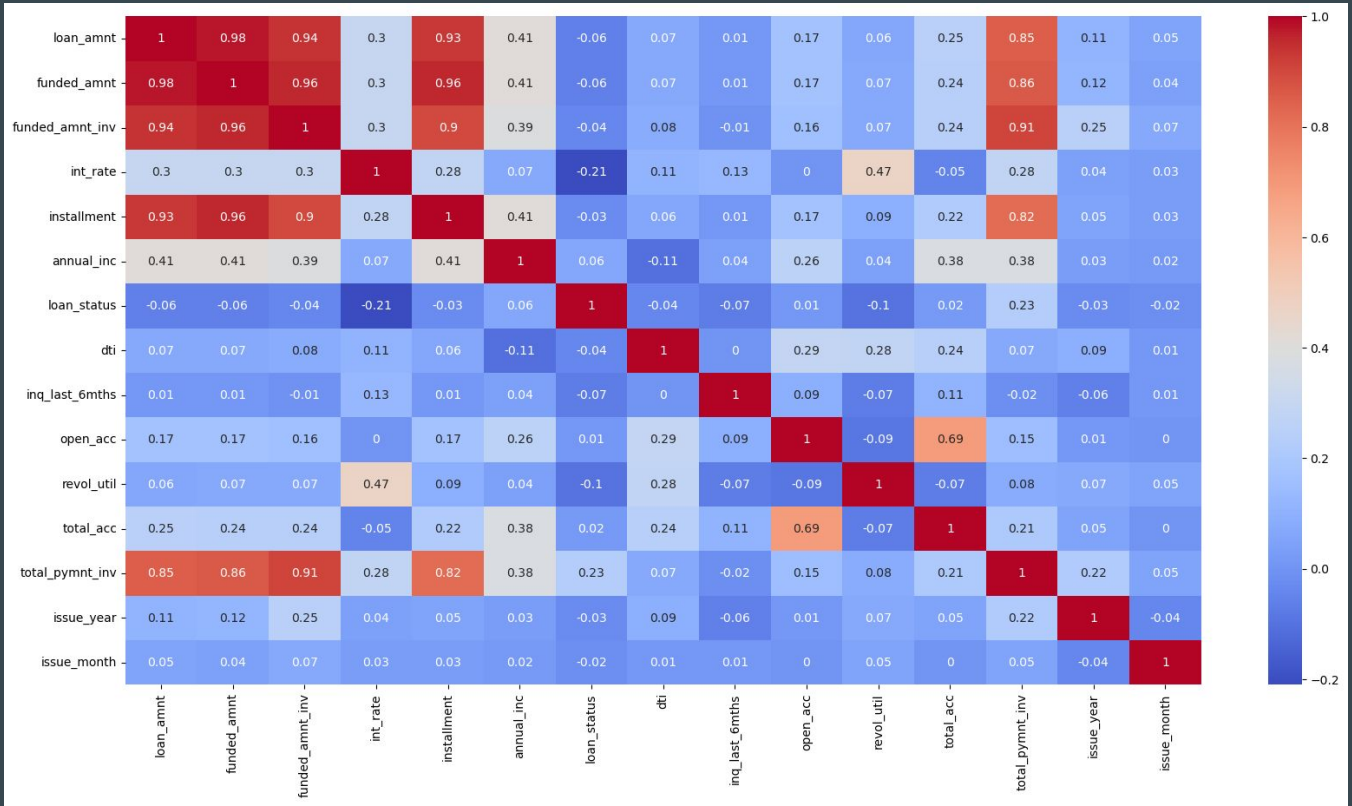




OBSERVATION SUMMARY - BIVARIATE ANALYSIS

- Customers are likely to default if the employment length is greater than 10+ years
- Customers with 'RENT' Category of home ownership are likely to default more
- When loan purpose is 'Small Businesses' the charged off customers are highest
- As the loan grade is increasing from A to G, the charged off customers proportion is increasing
- It's risky to offer even small amounts of loans to customers with verification status as 'Not Verified'
- As the Rate of Interest increases, loan defaulters increase.
- For a term of 36 months if the interest range is not between 7.5% to 13% then customers are likely to default
- For a term of 60 months, if the interest range is not between 11% to 17% then customers are likely to default
- Defaulters ratio is highest for loans provided in June and December
- Defaulters were highest in year 2011

Multivariate Analysis



Multivariate Summary

- Loan Status is highly negatively correlated with Revolving Line Utilization Rate (revol_util). This is quite obvious as the customer using most of their credit balance elsewhere are likely to default
- Loan Status is highly negatively correlated with Interest rate. The higher the interest rate, the higher the chances of Charge off
- Loan Status is highly positively correlated with total_pymnt_inv. This means those who already paid a major portion of their loan are like to fully pay it
- Loan Status is positively correlated with Annual Income. The higher the Income, the more likely the customer is going to pay off the loan

Conclusion

- Customers are likely to default if the employment length is greater than 10+ years
- Customers with 'RENT' Category of home ownership are likely to default more
- When loan purpose is 'Small Businesses' the charged off customers are highest
- As the loan grade is increasing from A to G, the charged off customers proportion is increasing
- It's risky to offer even small amounts of loans to customers with verification status as 'Not Verified'
- As the Rate of Interest increases, loan defaulters increase.
- For a term of 36 months if the interest range is not between 7.5% to 13% then customers are likely to default
- For a term of 60 months, if the interest range is not between 11% to 17% then customers are likely to default
- Defaulters ratio is highest for loans provided in June and December
- Defaulters were highest in year 2011
- The higher the Revolving Line Utilization Rate, the more likely the customers are going to default
- The higher the Annual Income, the less likely the customers are going to default