upGrad

Lending Club Case Study

C-53 – Batch 2023 – Submission Date 4th July 2023

Please find the details :

Group Facilitator : Name: SHRINIVAS NANDKISHOR DHAMONE
Email ID: shrinivas.d33@gmail.com
Phone No:919823352202

Team Member Detail: Name: Mohammed AbuSaleha
Email ID: mabusaleha@yahoo.com
Phone no: 919880118997

Lending Club Case Study

Problem Statement

Consumer finance company - lends various types of loans. For each loan application, make loan approval decision based on the applicant's profile.

Two types of risks:

```
- If the applicant is likely to repay the loan. Not approving == loss of business
- If the applicant is not likely to repay the loan == may lead to a financial loss for company(credit loss)
```

Data

Information about past loan applicants and whether they 'defaulted' or not.

Aim

Identify patterns which indicate if a person is likely to default. If one is able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. Identification of such applicants using EDA is the aim of this case study. In other words, the company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.

Action

Denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

Understand

Use EDA to understand how consumer attributes and loan attributes influence the tendency of default.

Scenarios

- Fully paid: The principal and the interest rate are fully paid.

- Current: In the process of paying the instalments & the loan tenure not yet completed. Not 'defaulted'.

- Charged-off: Not paid the instalments in due time for a long period of time. Defaulted on the loan

Who are 'defaulters'. ?

The customers labelled as 'charged-off' are the defaulters

upGrad

Index

- EDA approach
- Uni-variate Analysis
- Bi-Variate Analysis
- Co-relation Analysis
- Final Summary / Conclusion

EDA

- Checked info, describe, shape for given Data.
- Checked Data dictionary for Business Lingo
- Cleaned up unwanted columns (containing all zeros / NaN's)
- Identified and Treated outliers with function and IQR method
- Removed columns not relevant to this case study
- Created some derived columns to Category/ Bin of Annual Income, Interest rates etc.
- Identified Continuous and Categorial variable.
- Once Data is cleaned, performed Univariate, Bivariate and Multi Variate analysis.
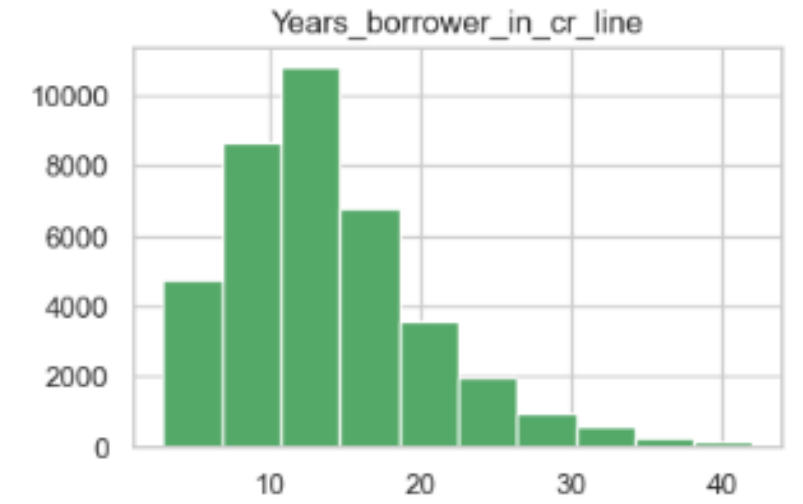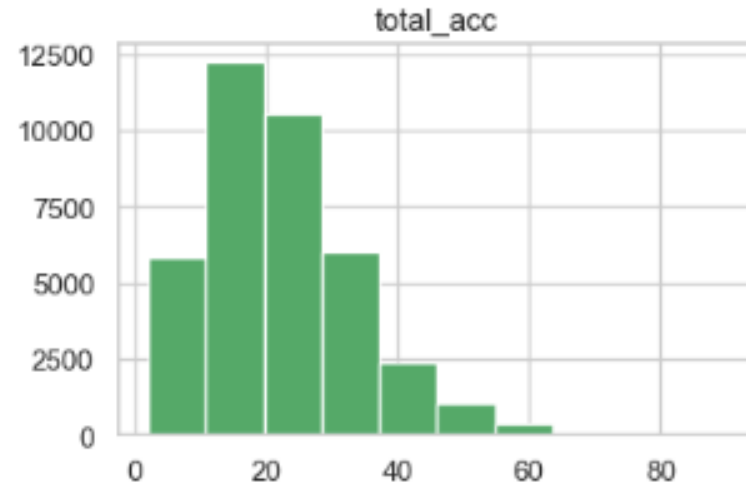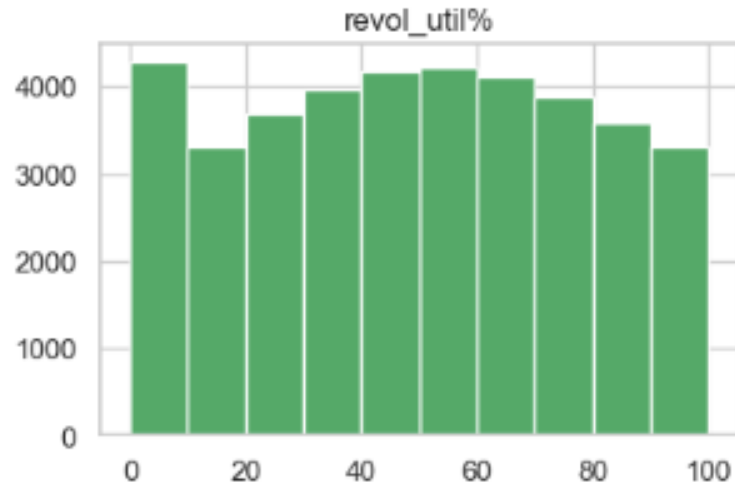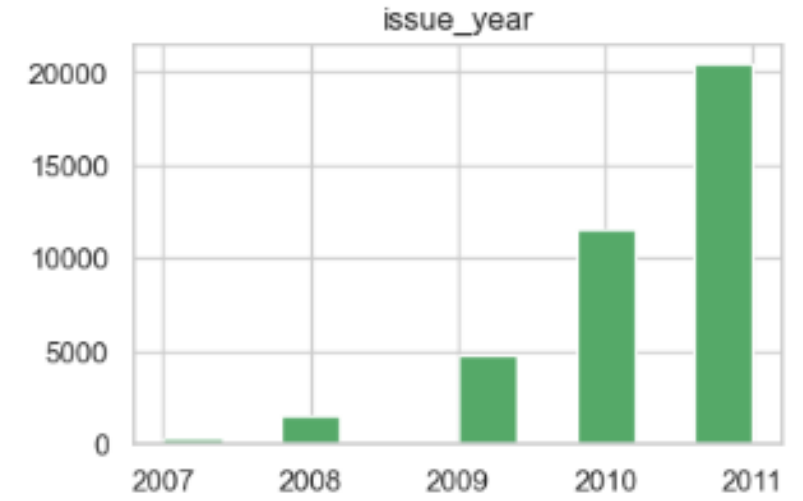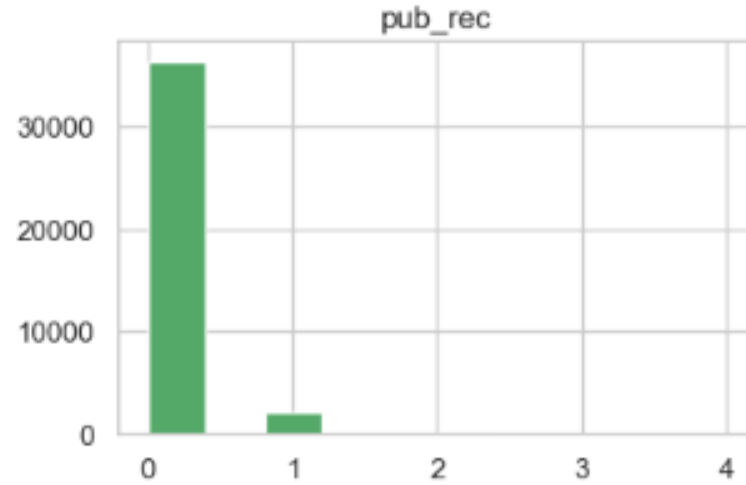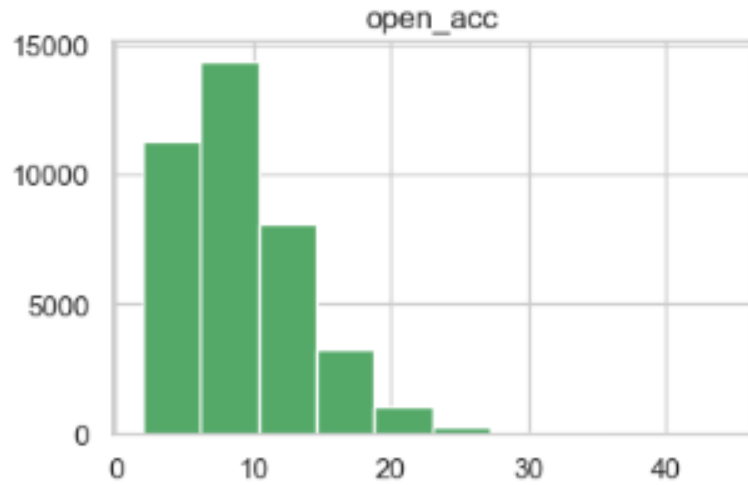- Summarized the outcome of this case study

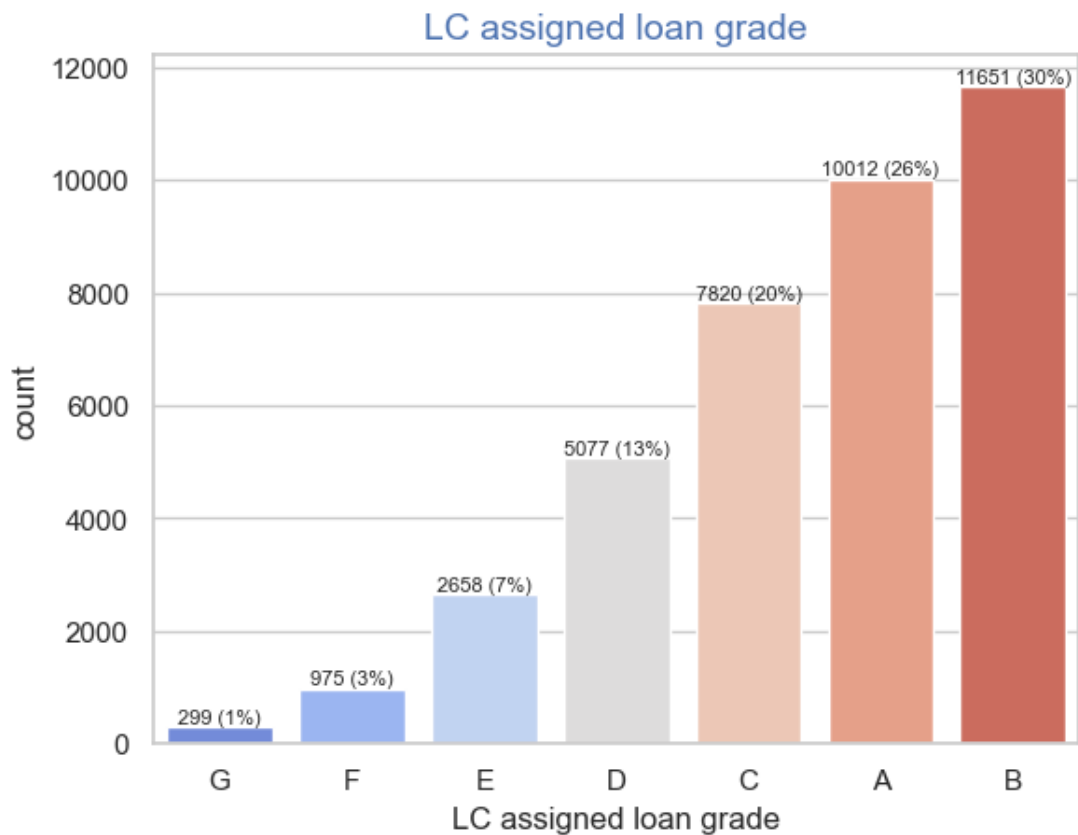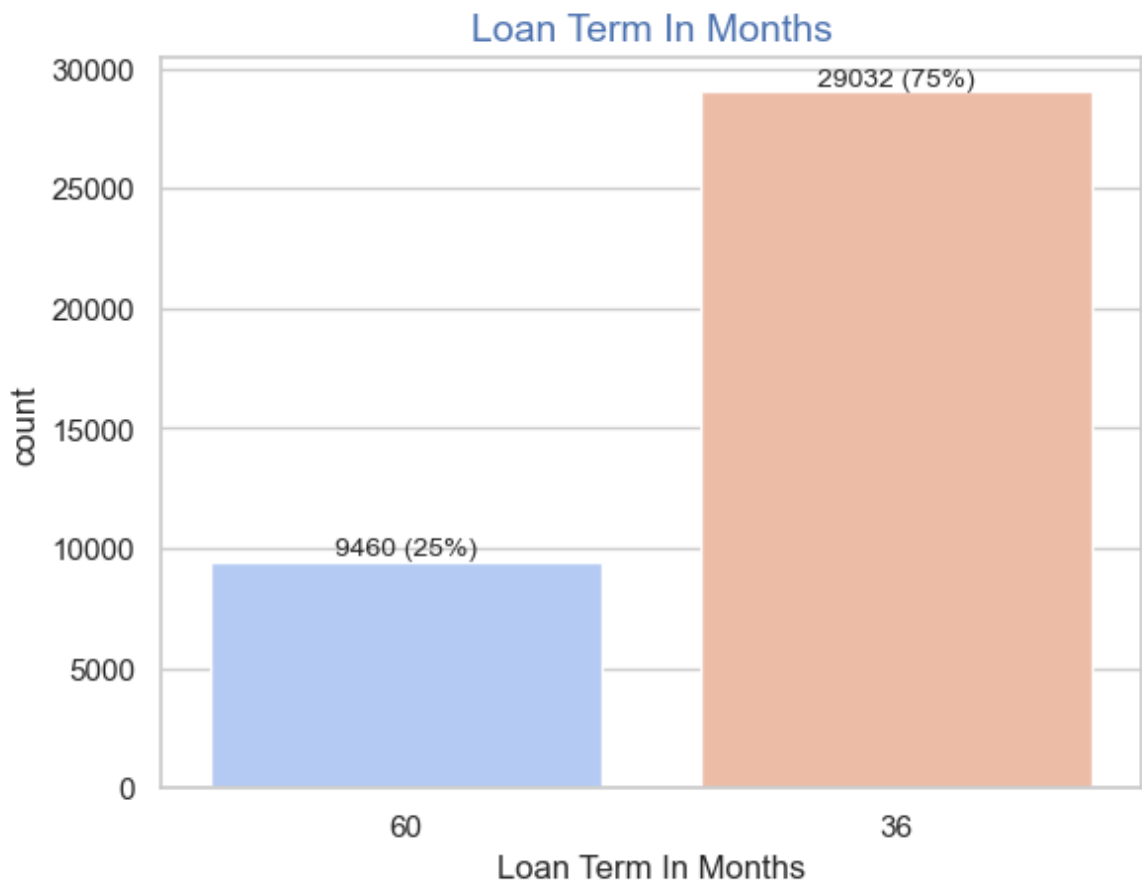# Assumptions

Ignored "Current/ ongoing loans" Status

Data Distribution – quick Glance - Important Factors
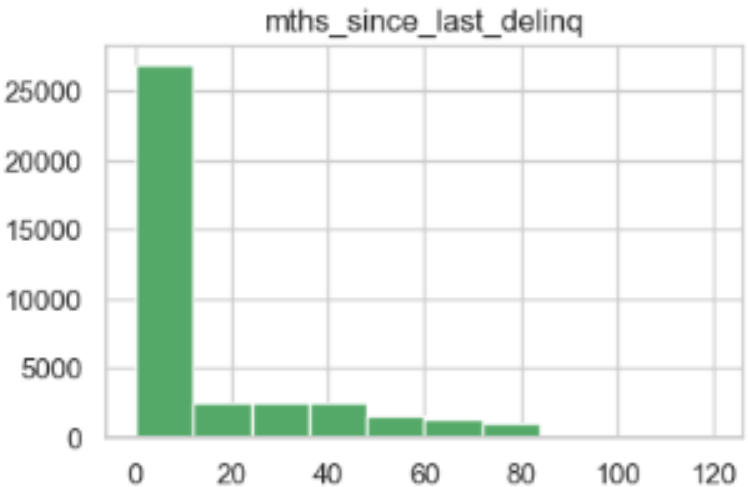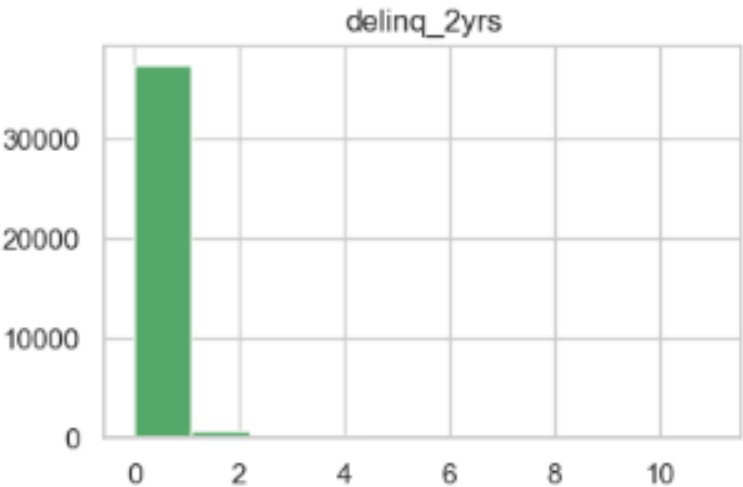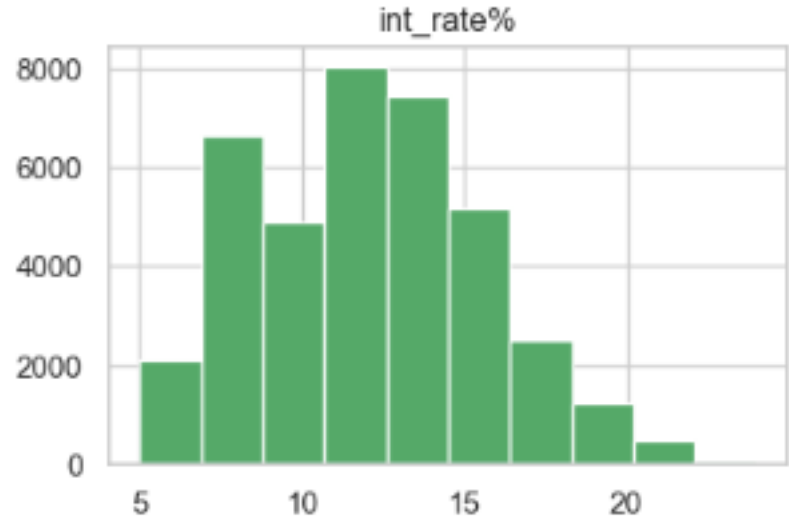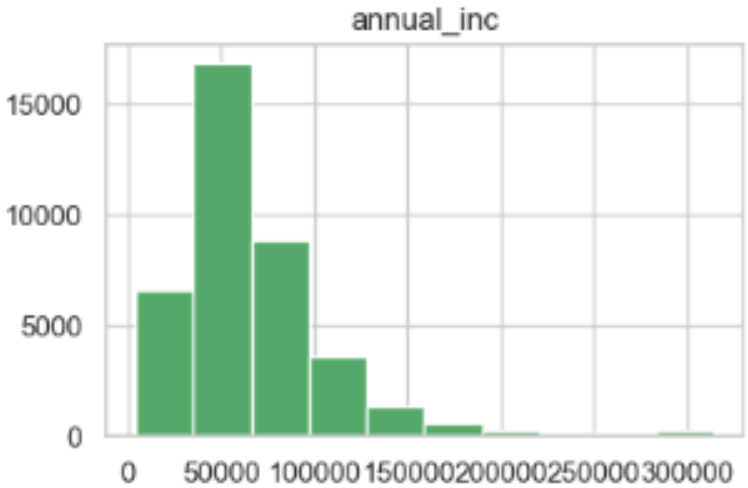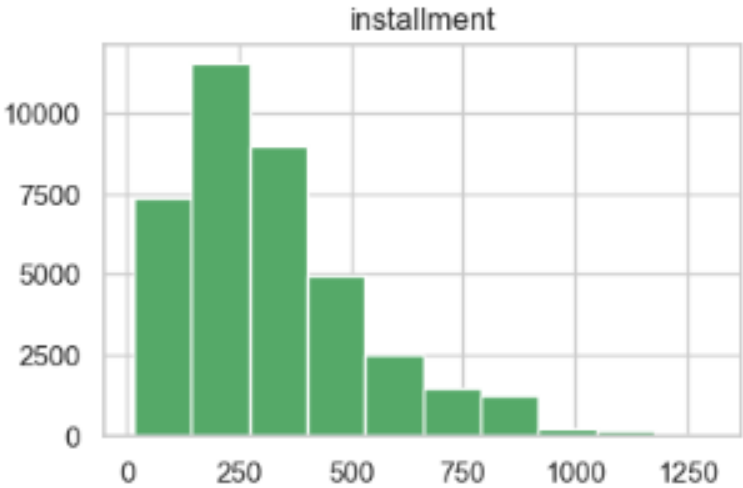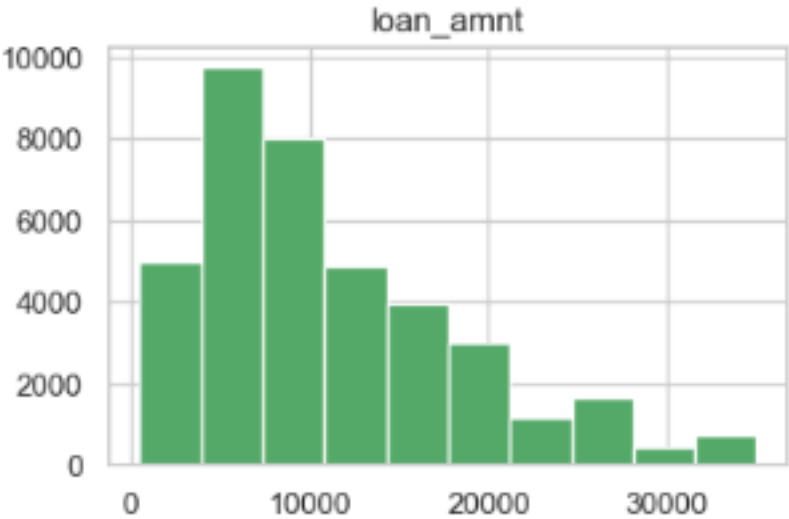
Data Distribution – quick Glance - Important Factors

Univariate Analysis – Few important Columns

Observations: We have more cases of 36 months loan - 75%

Observations: We have more of A(26%) and B (30%) Loan Grades

Univariate Analysis

Employee Working Experience in Years
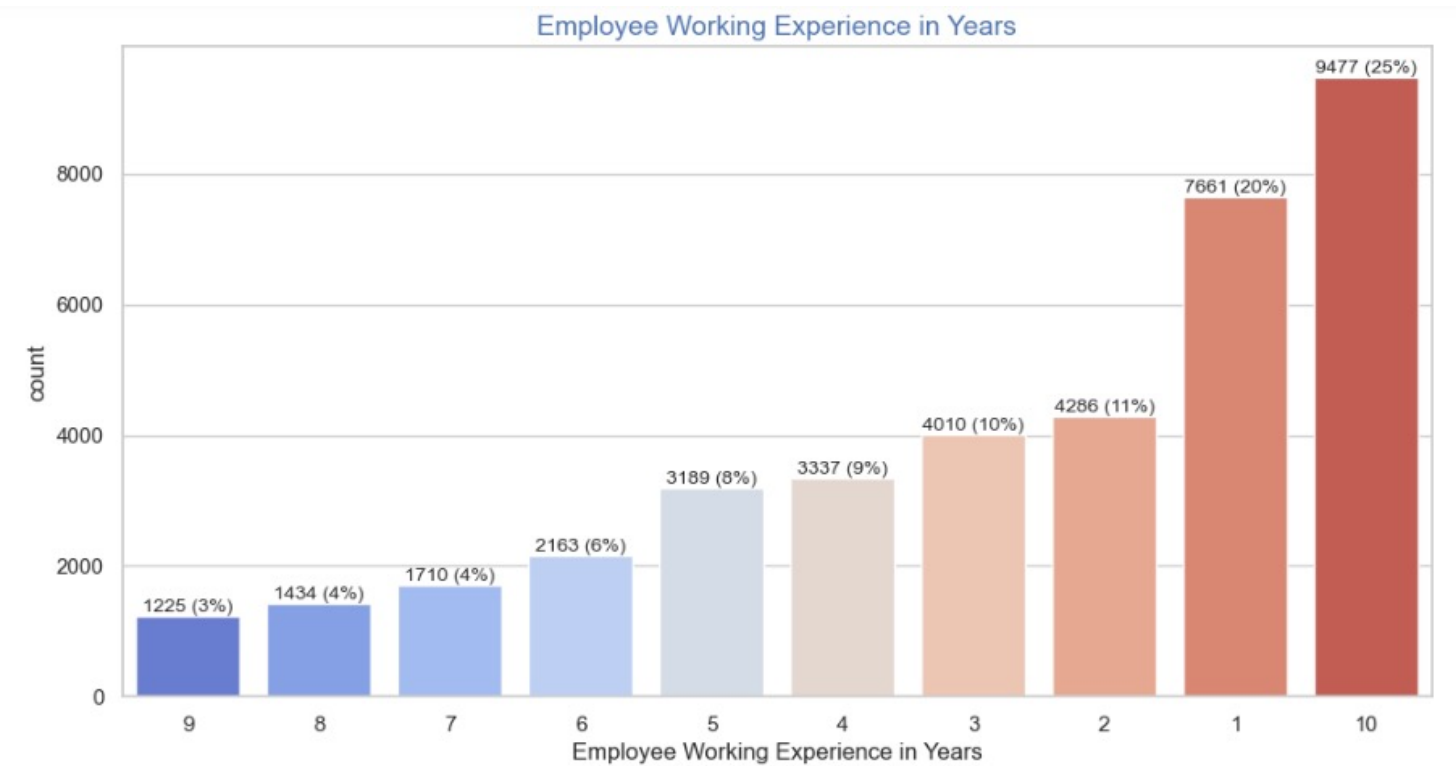
Observations: We have more of 10 years(25%) and 1 year(20%) employee work experience

Borrower Home Ownership
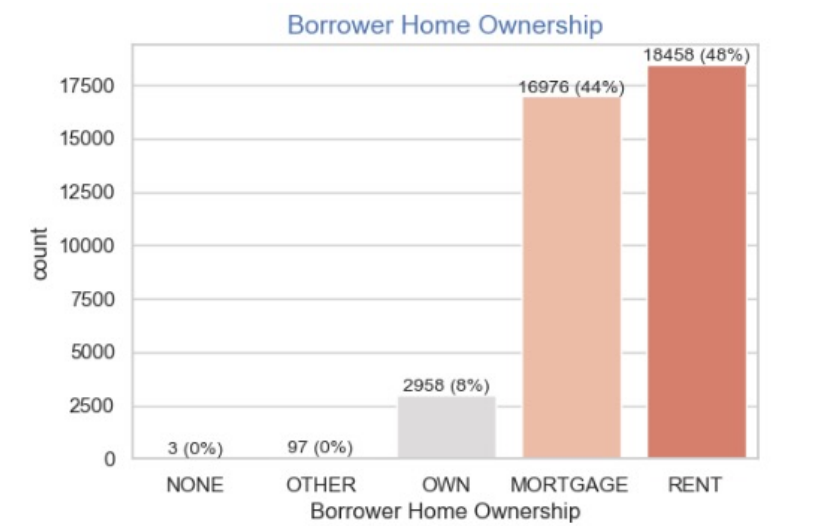
Observations: We have more of Rent(48%) and Mortgage(44%) type home ownership

Income Verified by LC
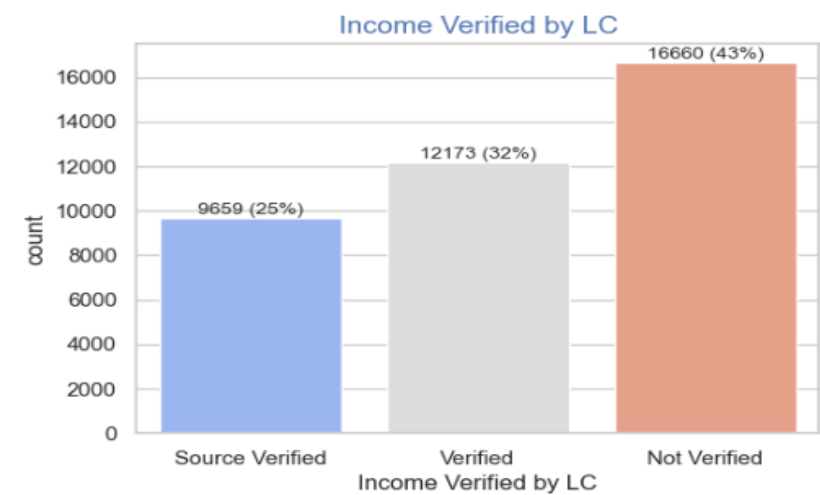
Observations: We have more of Not_Verified(43%) as Verification status

Univariate Analysis

Current status of the loan

- Charged Off: 5611 (15%)
- Fully Paid: 32881 (85%)
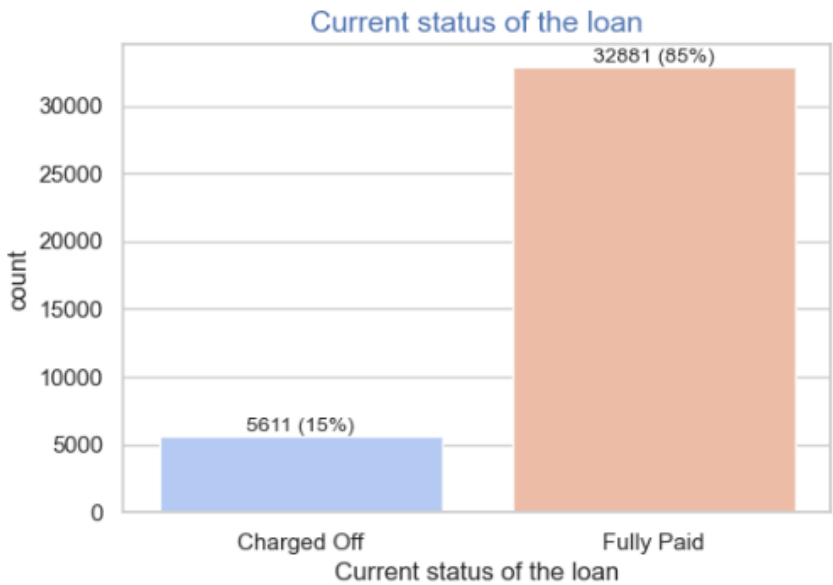
Observations: We have more of Fully Paid cases - 85%

Loan Amount

- 28000+: 1137 (3%)
- 21000-28000: 2783 (7%)
- 14000-21000: 7023 (18%)
- 7000-14000: 13278 (34%)
- 0-7000: 14271 (37%)
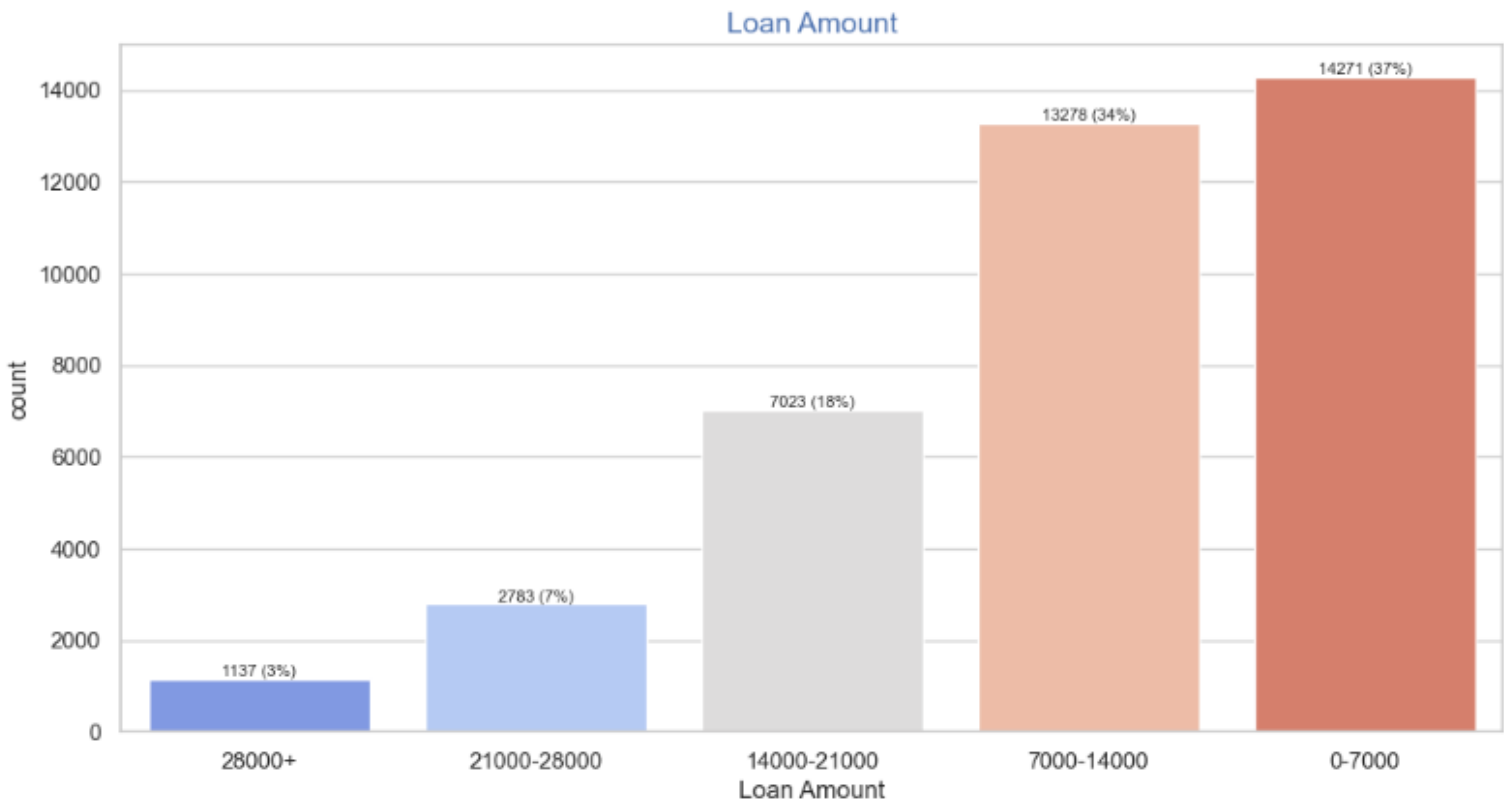
Observations: We have more of Loans of 0-7000(37%) and 7000-14000(34%)

Univariate Analysis

Loan purpose
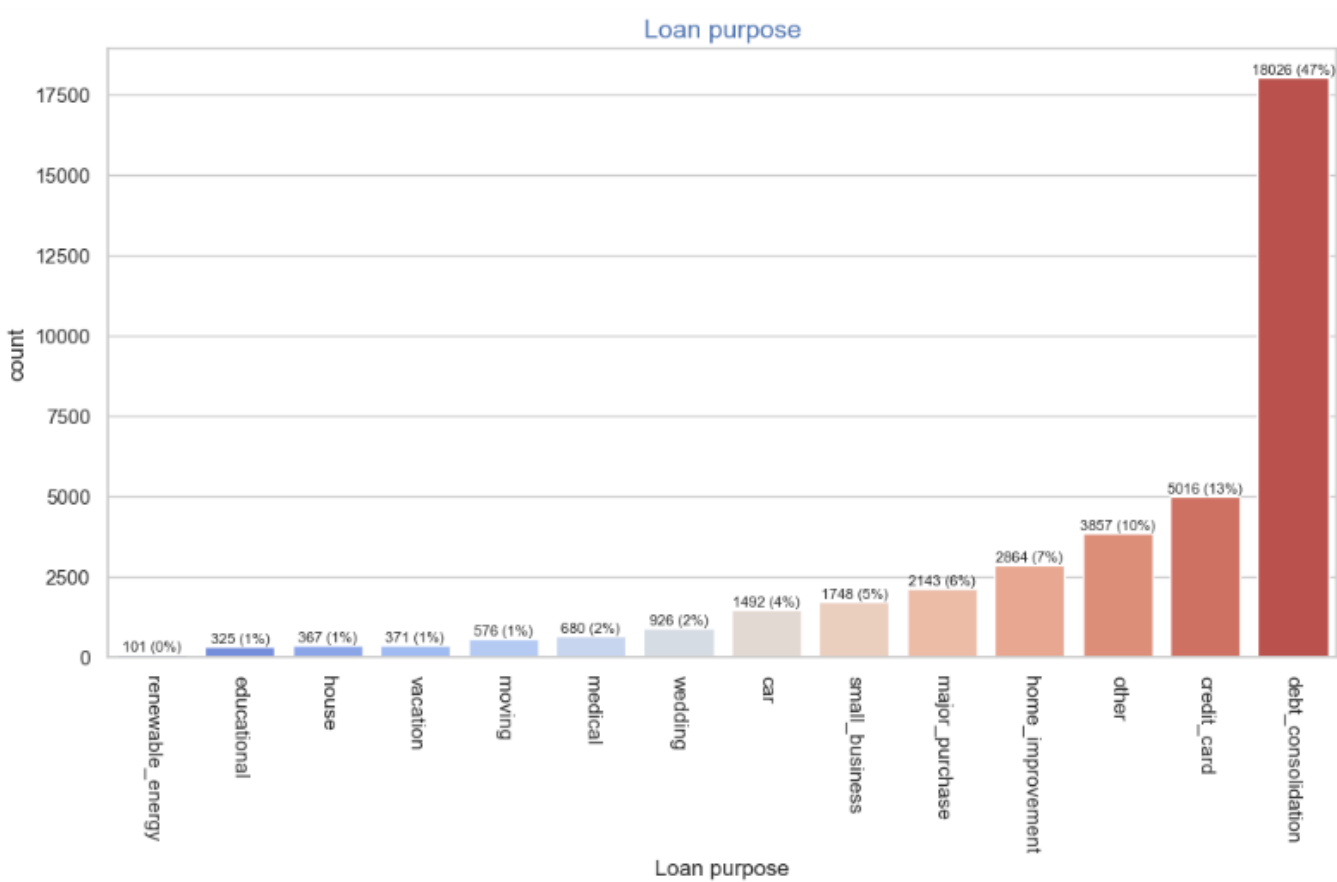
Observations: We have more of Debt Consolidation 47% & Credit Card 13%
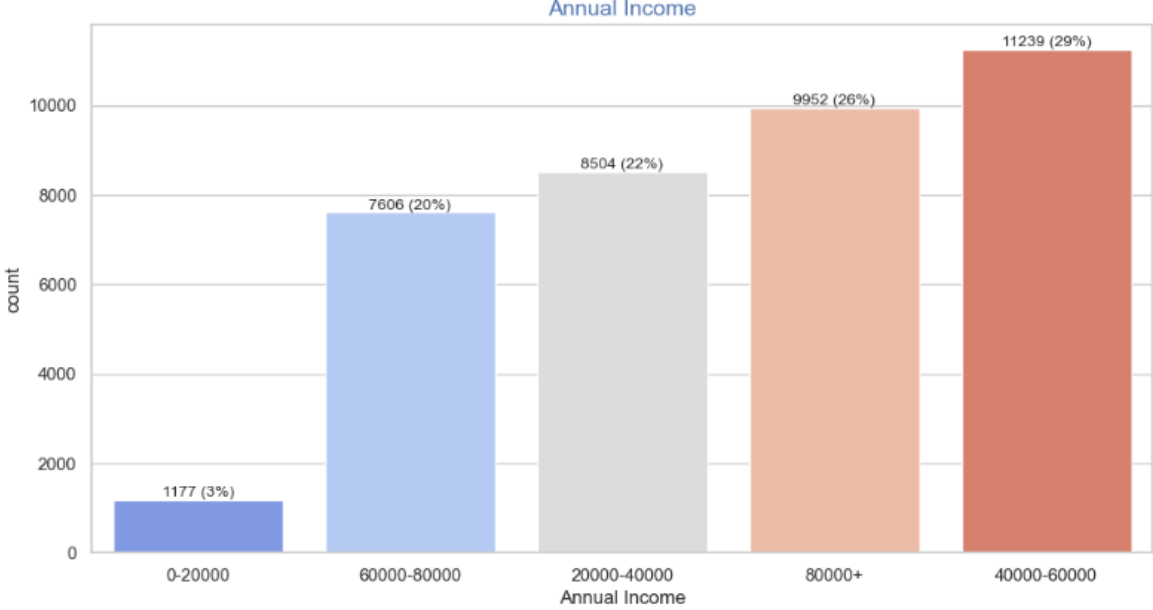
Annual Income

Observations: We have more of Annual Income of 40-60k(29%) and 80k+(26%)

Debt-to-income (DTI) ratio

Observations: We have more of dti of 10-15(25%) and 15-20(23%)

Univariate Analysis

Borrower Address State in the loan application
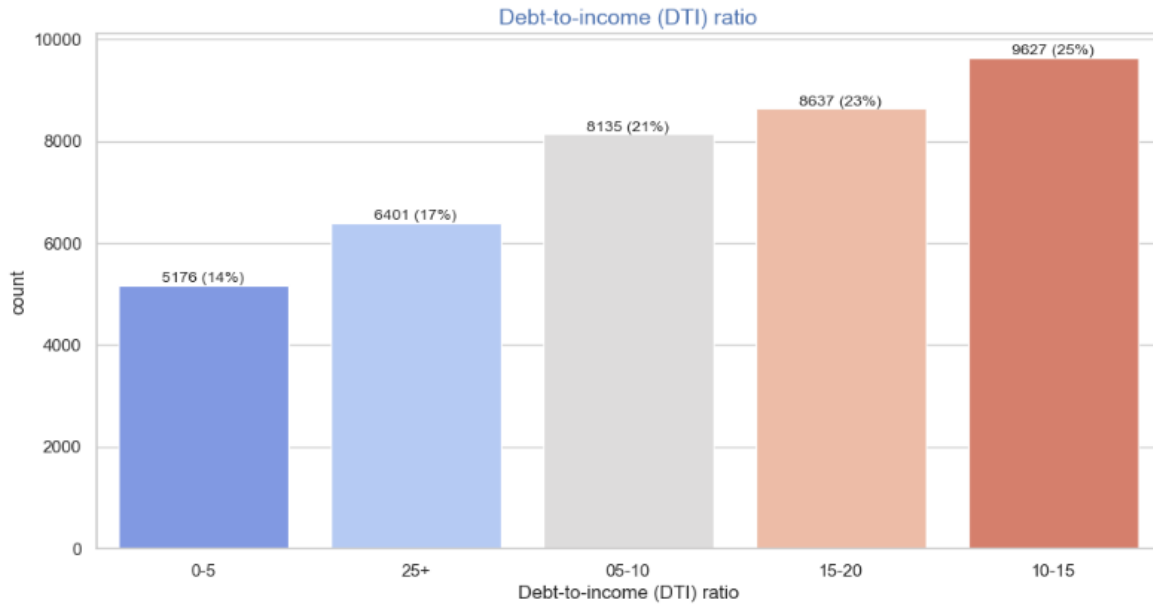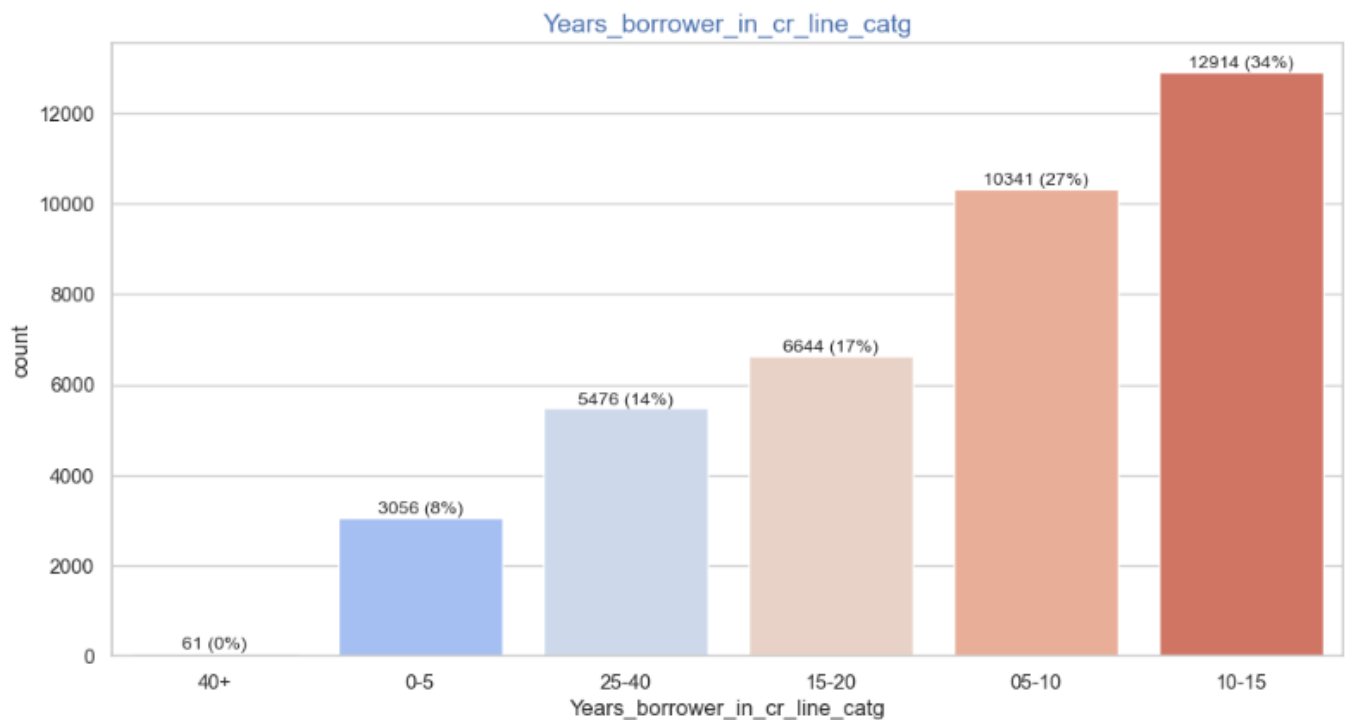
State	Count
CA	6929 (18%)
NY	3692 (10%)
FL	2772 (7%)
TX	2653 (7%)
NJ	1789 (5%)
IL	1476 (4%)
PA	1461 (4%)
VA	1367 (4%)
GA	1358 (4%)
MA	1294 (3%)
OH	1177 (3%)
MD	1022 (3%)
AZ	849 (2%)
WA	816 (2%)
CO	765 (2%)
NC	749 (2%)
CT	726 (2%)
MI	701 (2%)
MO	663 (2%)
MN	605 (2%)
NV	479 (1%)
SC	458 (1%)
WI	438 (1%)
AL	435 (1%)
OR	435 (1%)
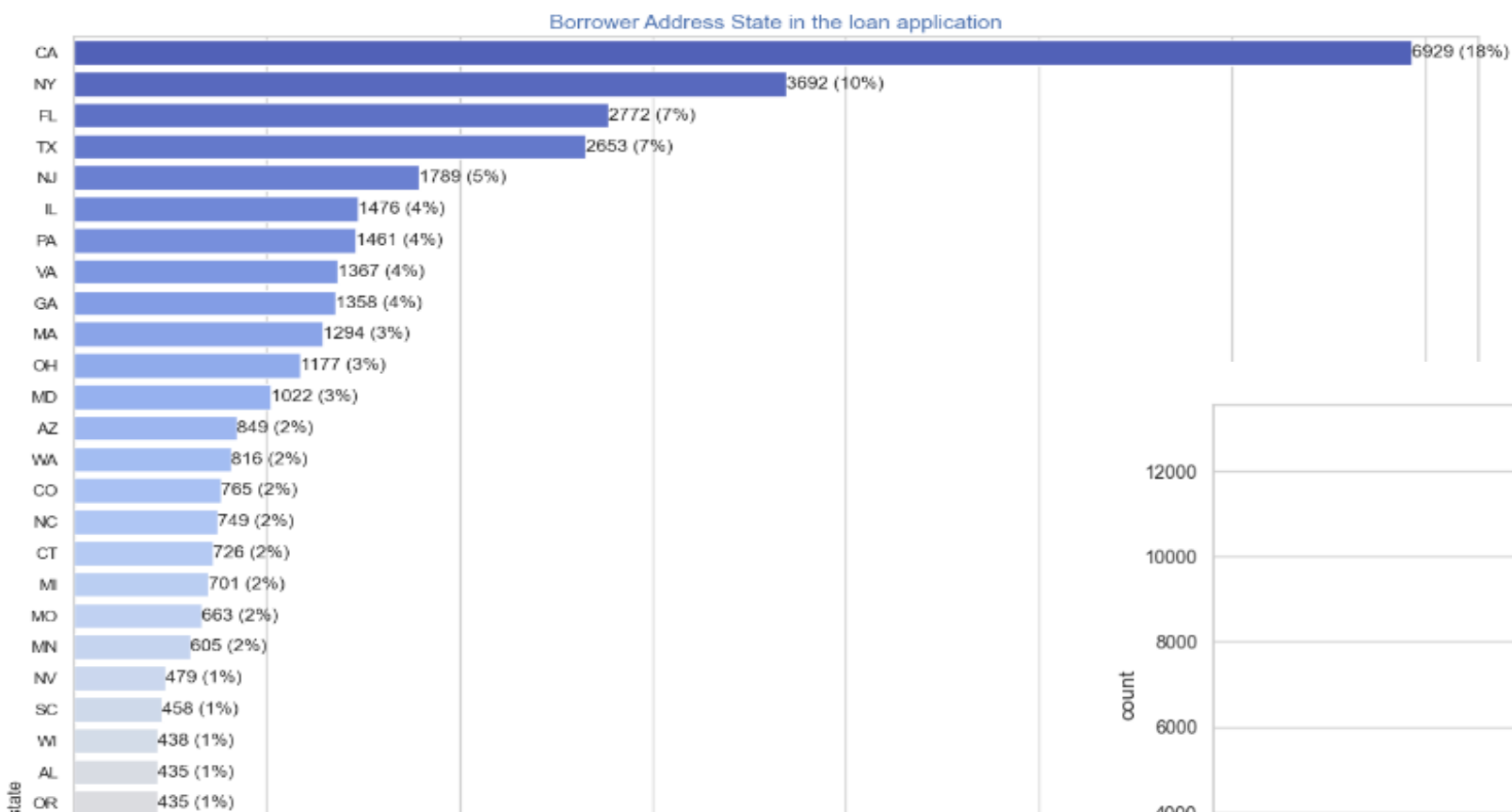
Observations: We have more of Loans from CA(18%) and NY(10%)

Years_borrower_in_cr_line_catg

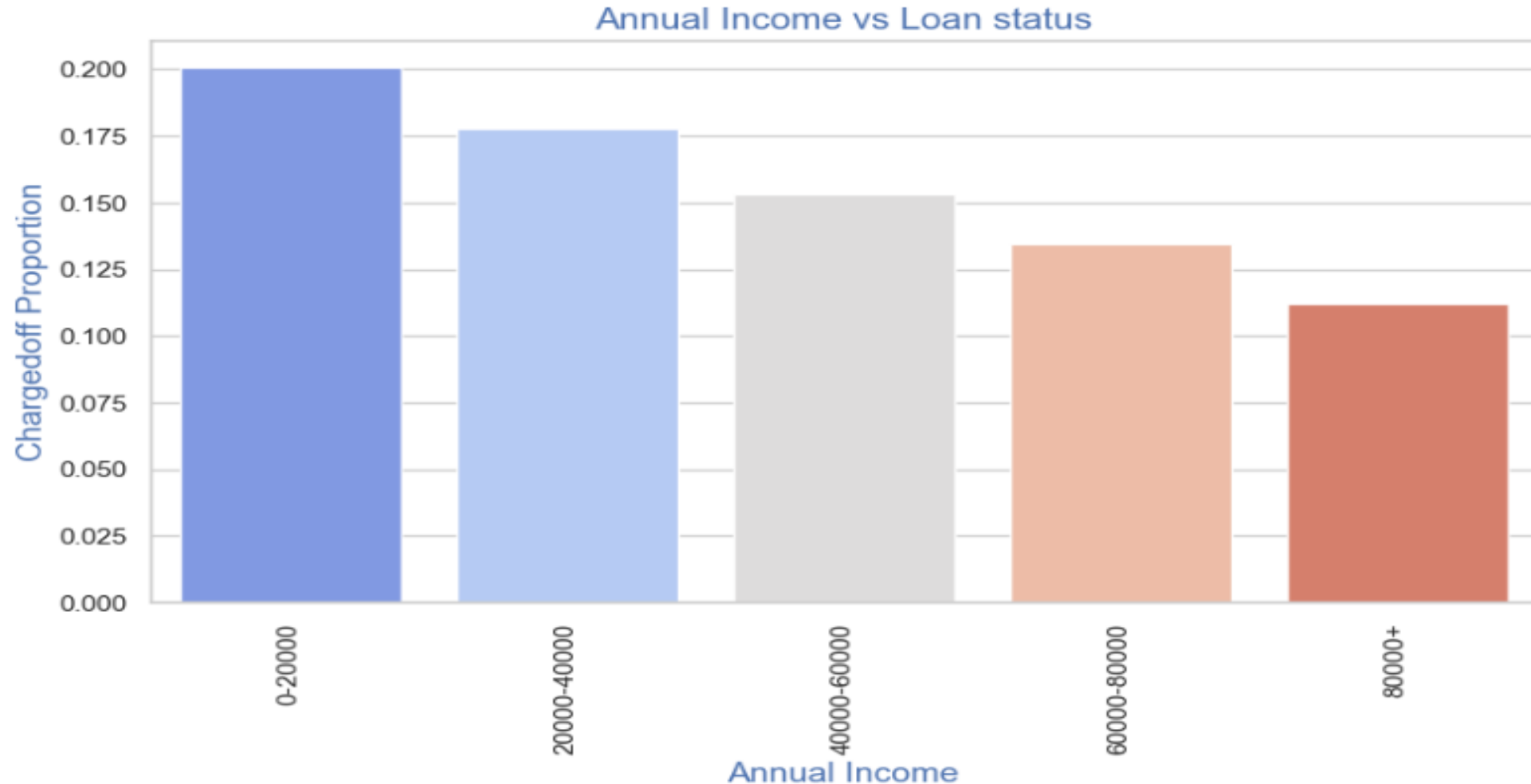| Years_borrower_in_cr_line_catg	count
40+	61 (0%)
0-5	3056 (8%)
25-40	5476 (14%)
15-20	6644 (17%)
05-10	10341 (27%)
10-15	12914 (34%)

Observations: We have more of borrower with credit history of 10-15 years(34%) and 5-10(27%)
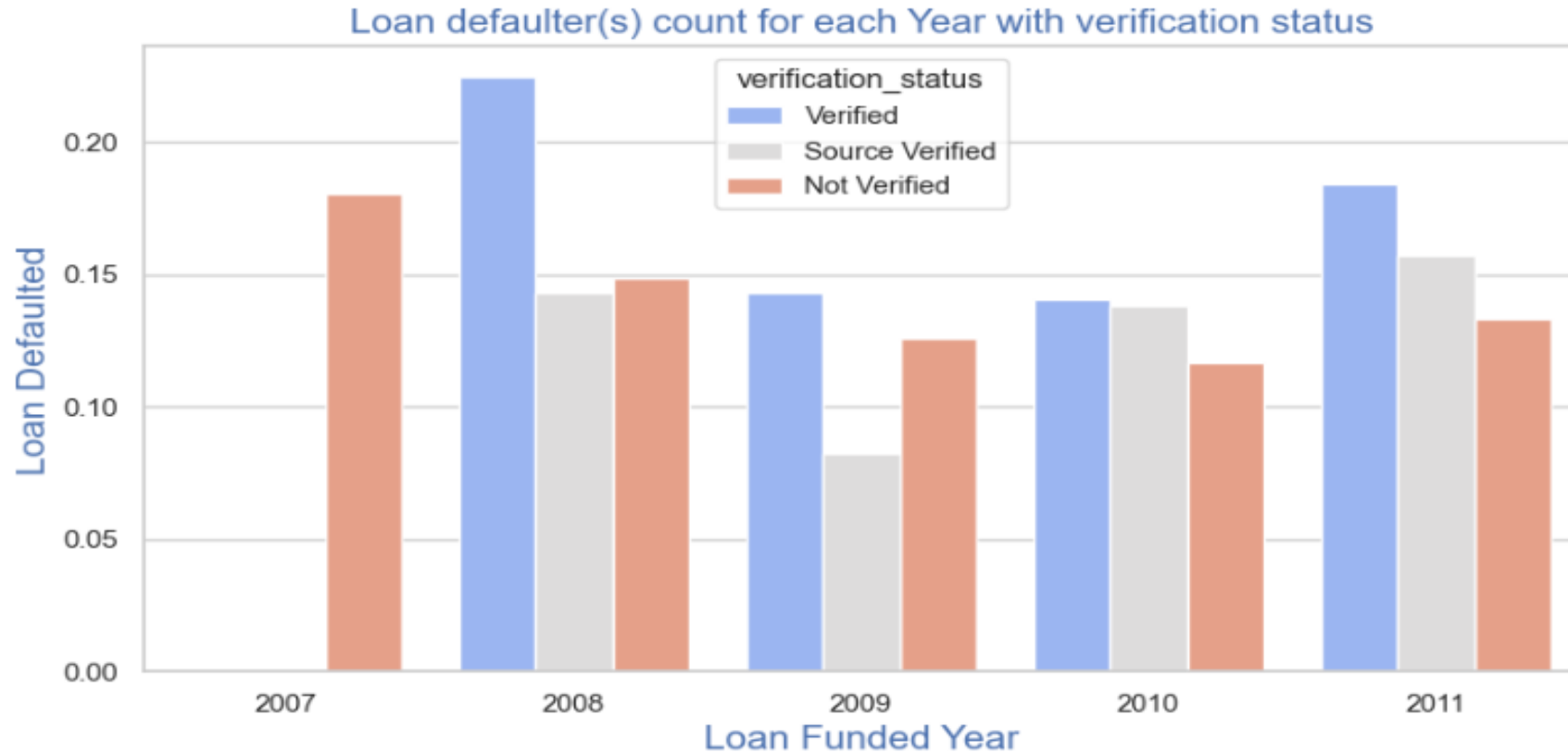
Univariate Analysis
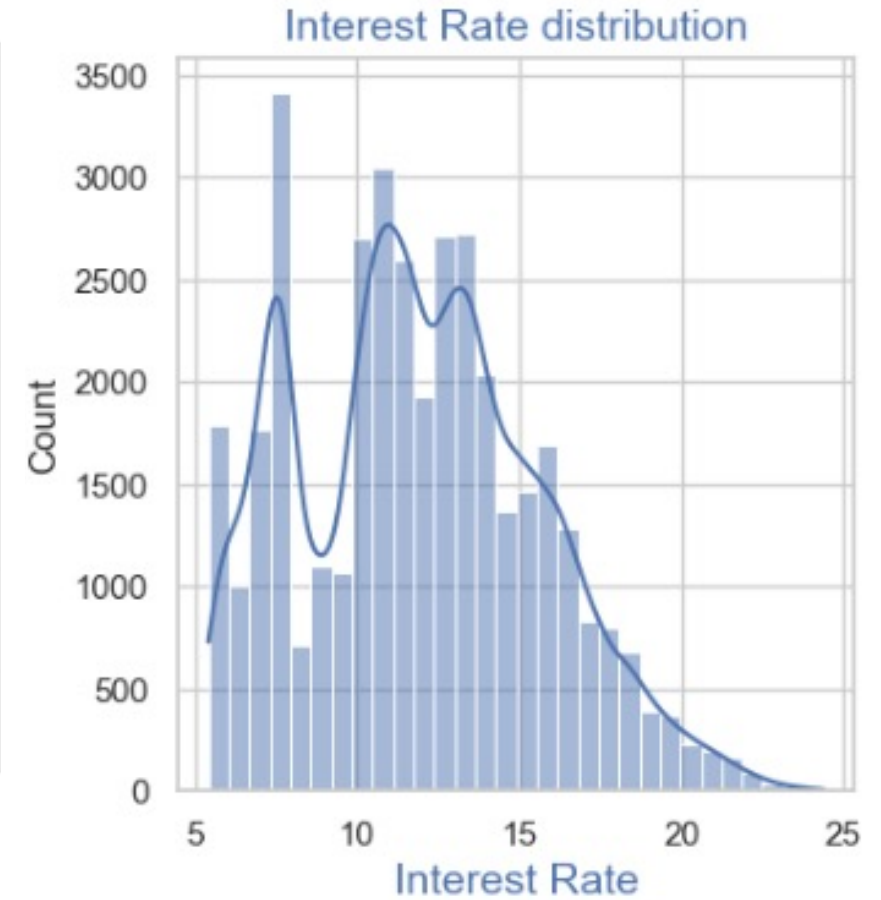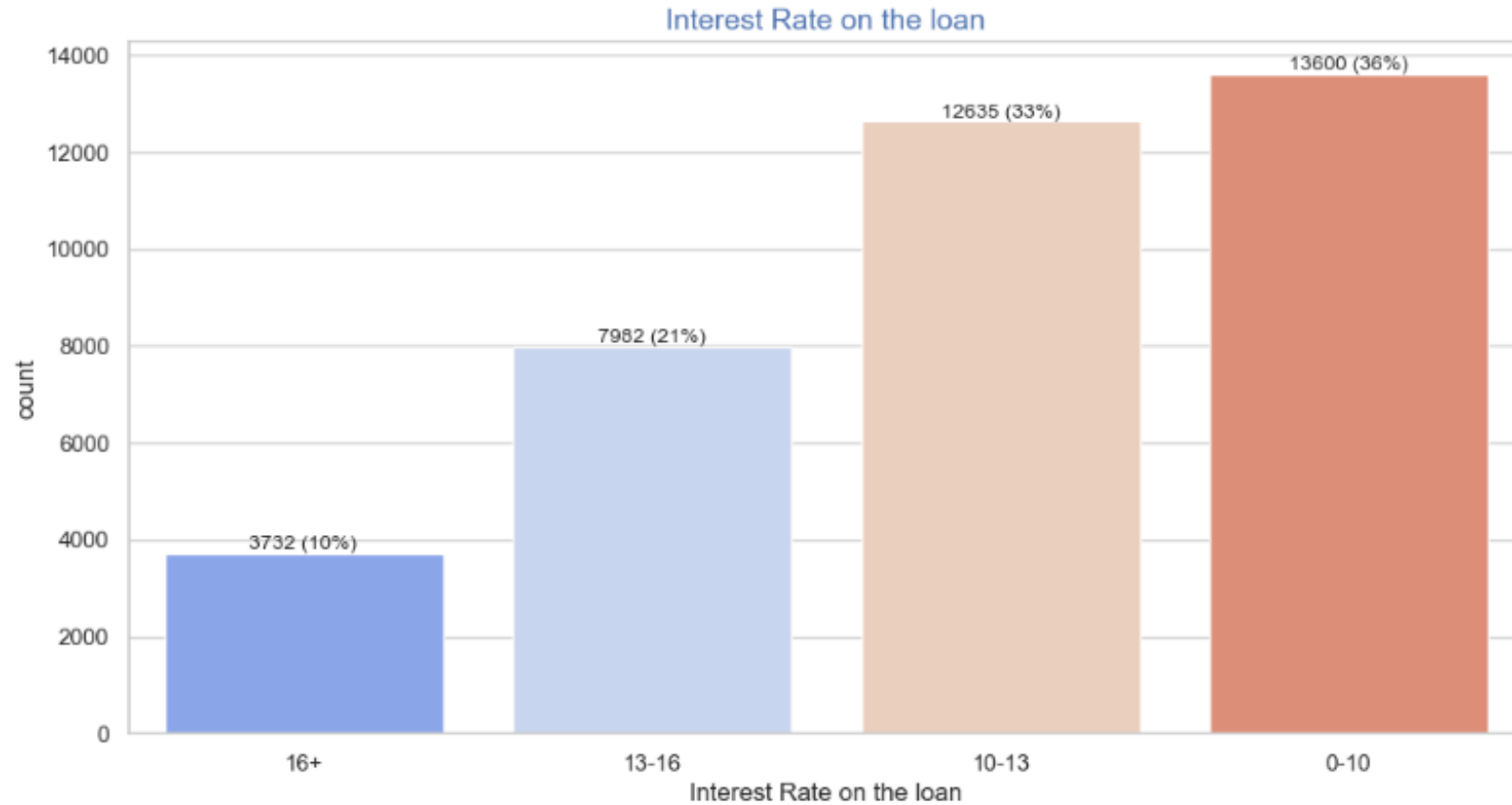
Observations: We have more of Interest Rate of 0-10%(36%) and 10-13%(33%)

Bi-variate Analysis – Few important Columns

Annual Income vs Loan status

Observation: Borrower with low income seens to have more Loan Defaulter(s)

Bi-variate Analysis – Few important Columns

Loan defaulter(s) count for each Year with verification status

Observations: Increasing number of defaulters over the year 2008. Number of defaulters have increased during 2008 - may be due to Global economy recession.

Bi-variate Analysis – Few important Columns
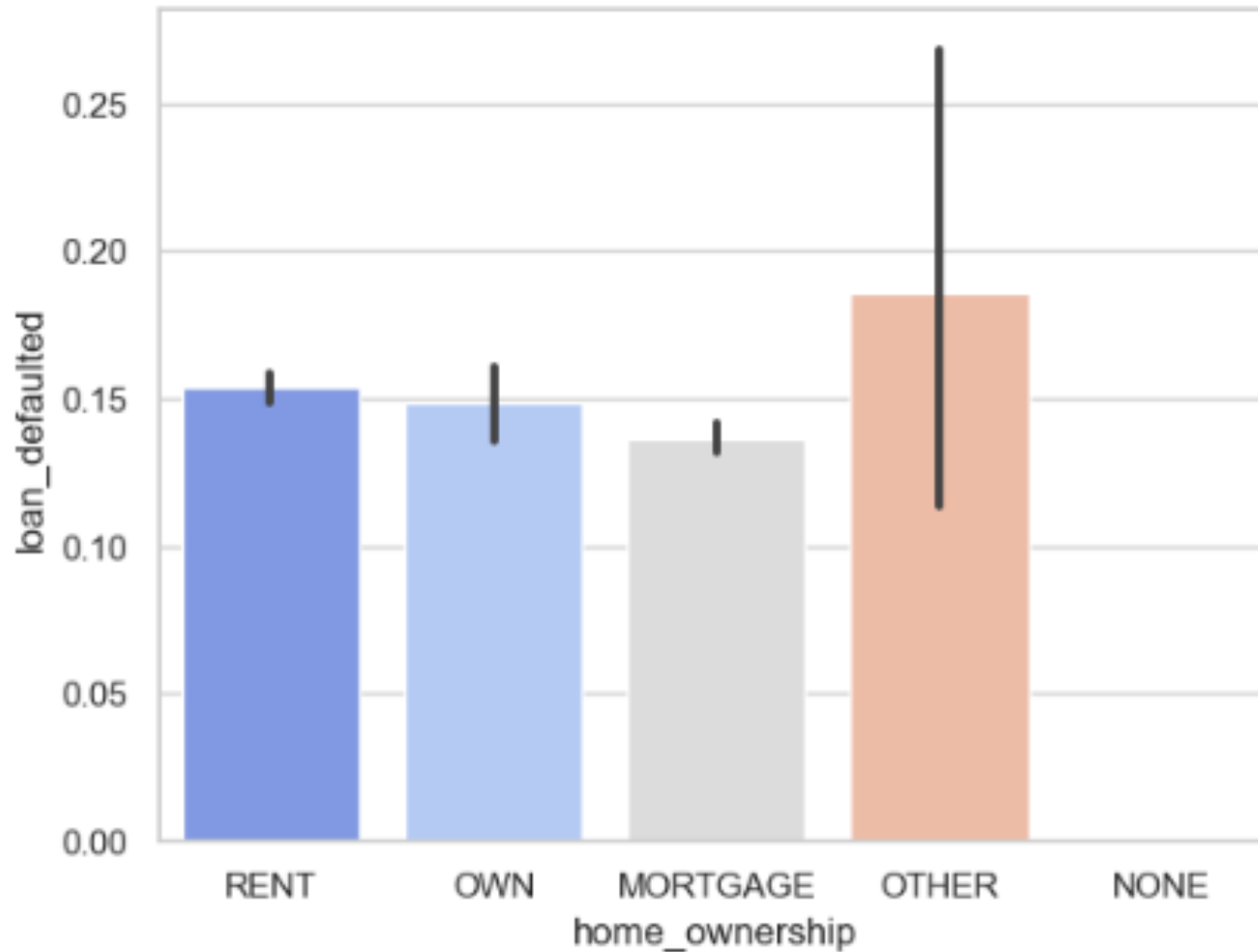
Observations: Other home-ownership have high chances of loan default

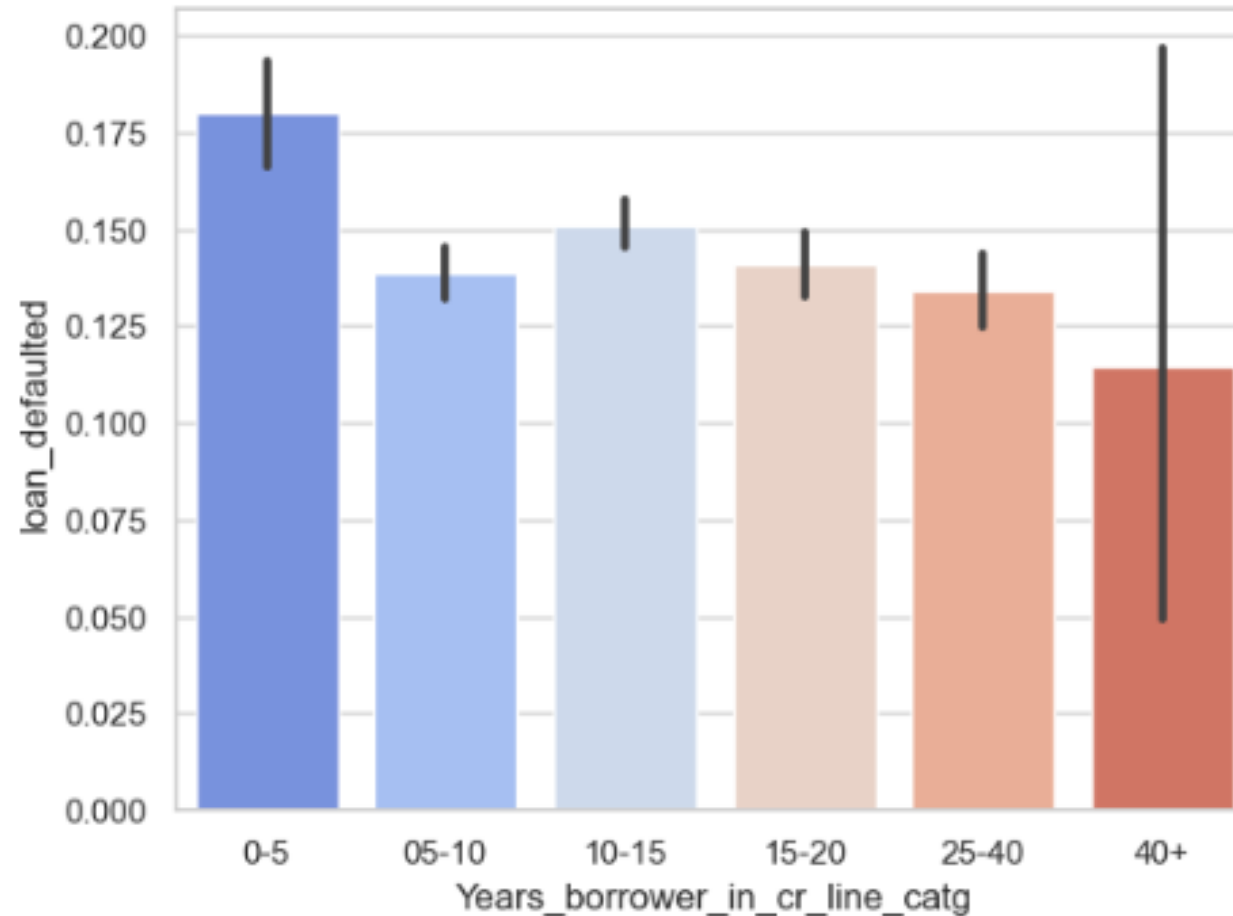# Bi-variate Analysis – Few important Columns

Observations: New Borrowers (0-5) and old borrowers (40+)

seems to have more loan defaults

Bi-variate Analysis – Few important Columns
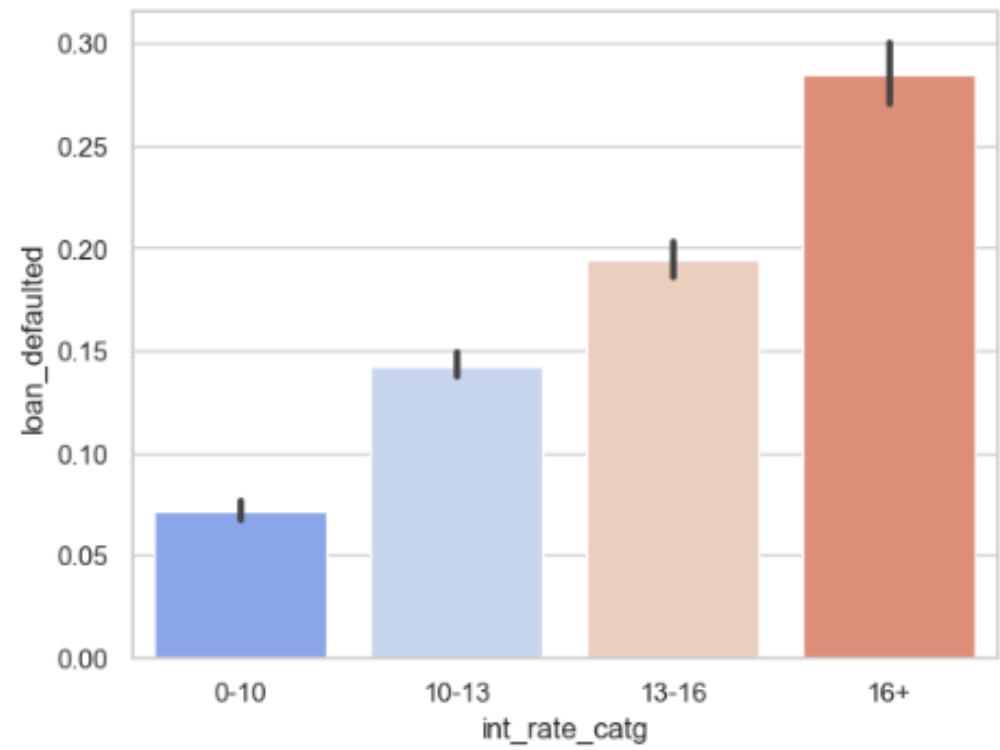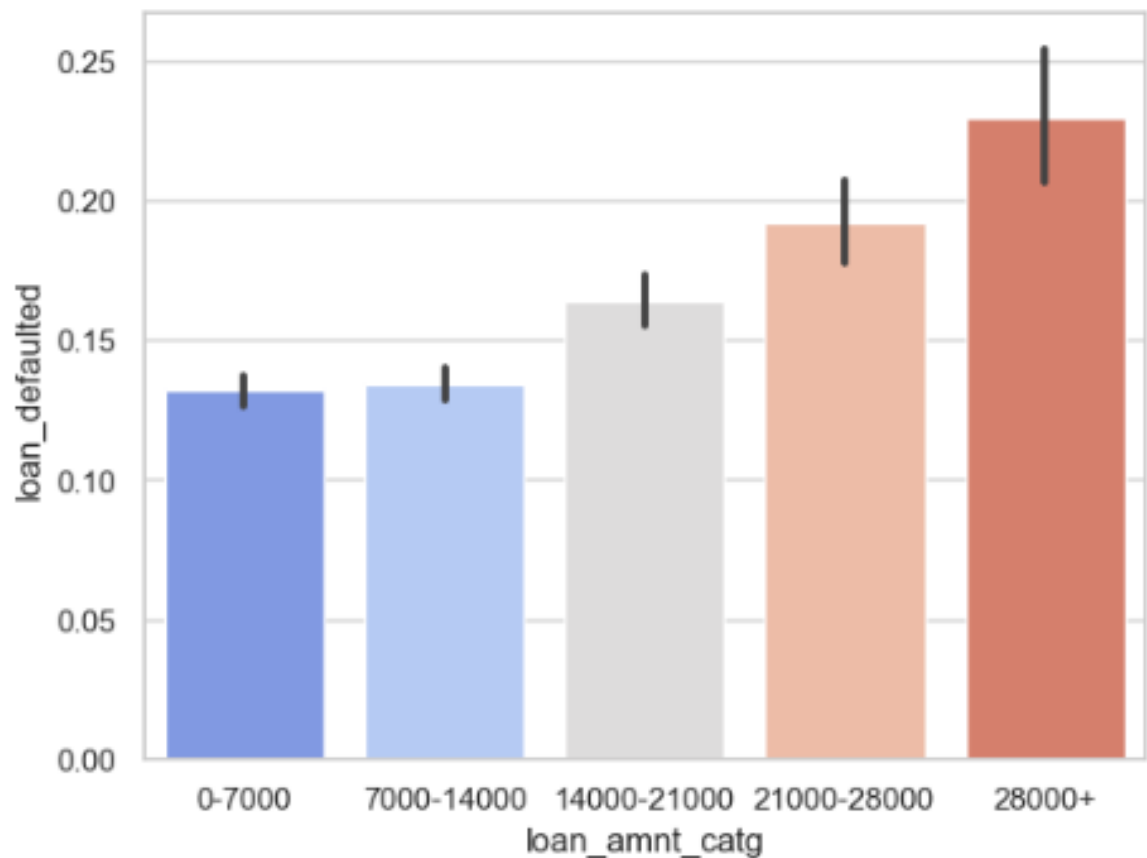
Observations: Higher interest rates seems to have higher loan defaulters

Observations: Higher loan amount have high chances of loan default

Bi-variate Analysis – Few important Columns
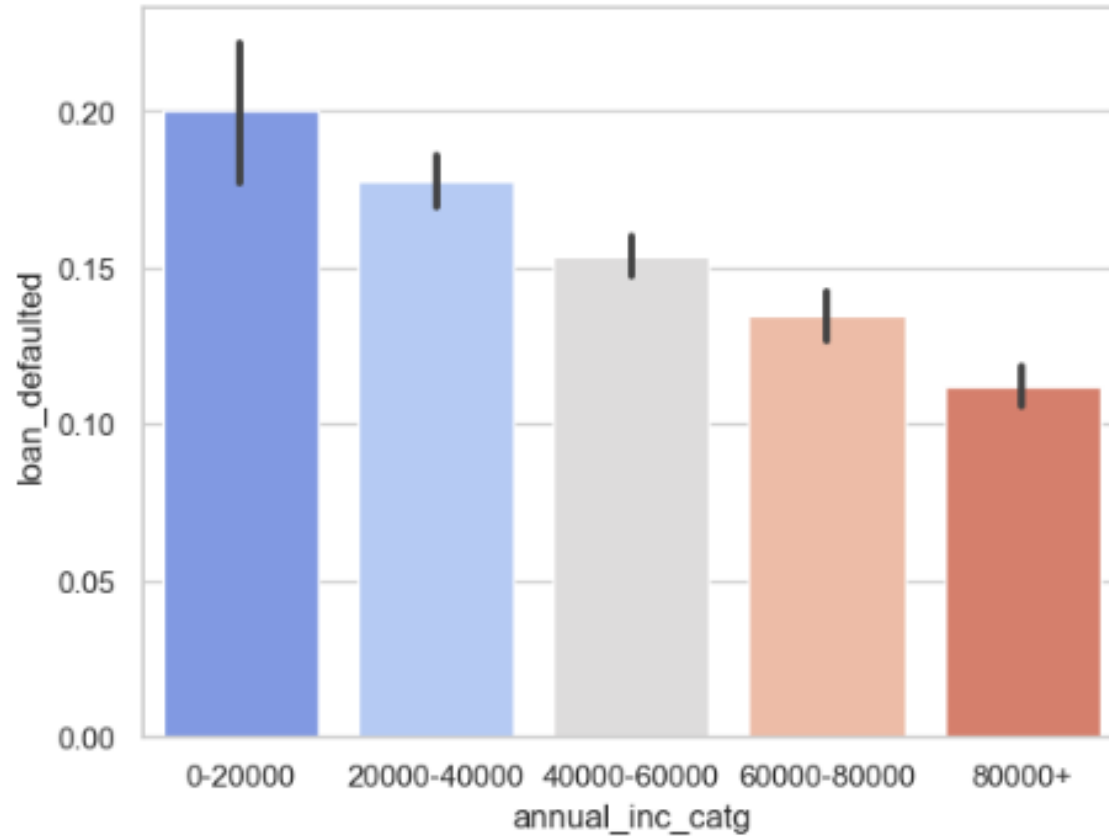
Observations: Borrower's with less income (0-20000) seems to default more

Bi-variate Analysis – Few important Columns

Purpose of Loan vs Loan Amount
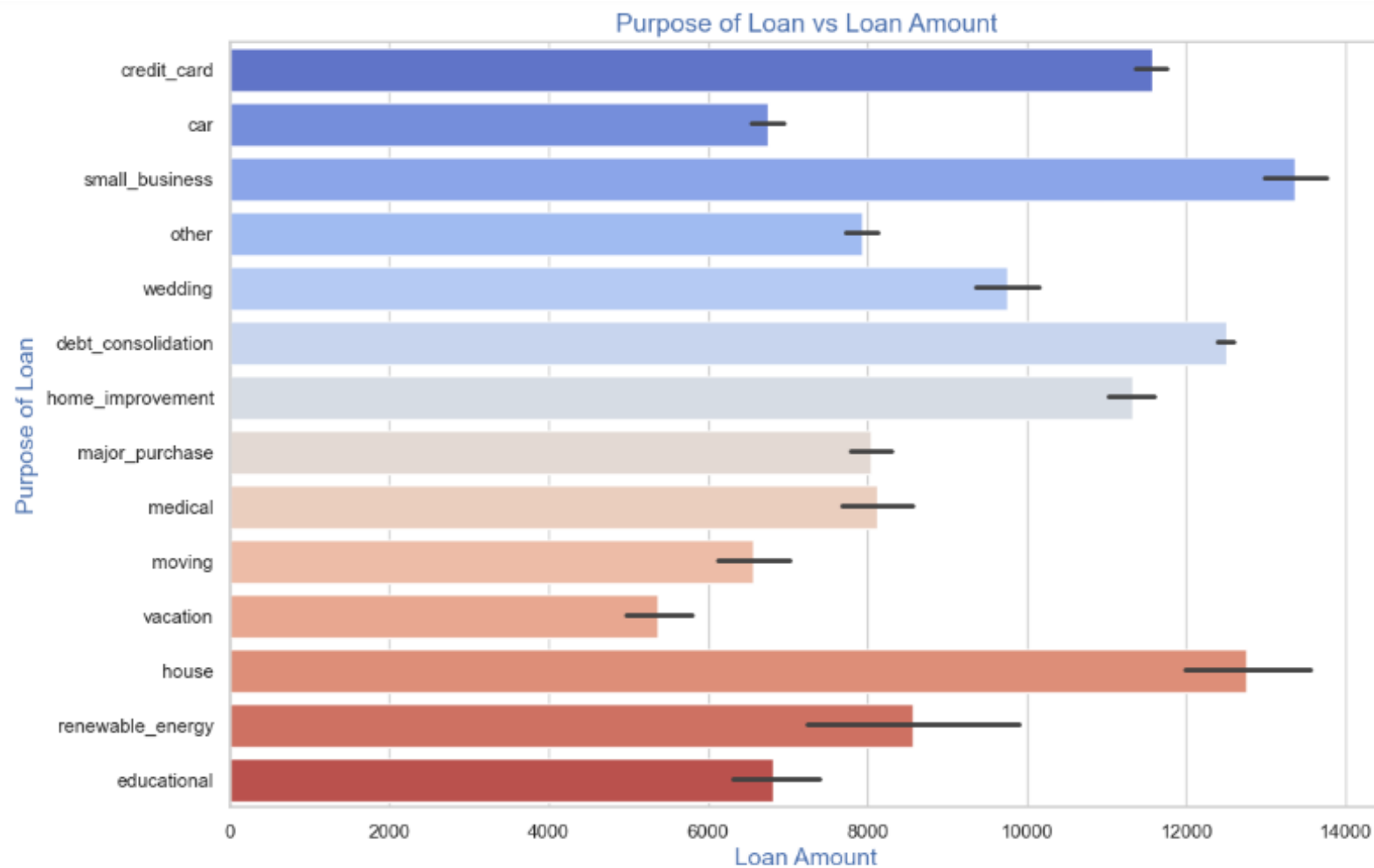
Observations: Small Business applicants have high chances of getting charged off as their revenue generation may be low.

Bi-variate Analysis – Few important Columns

Grades vs Chargedoff Proportion
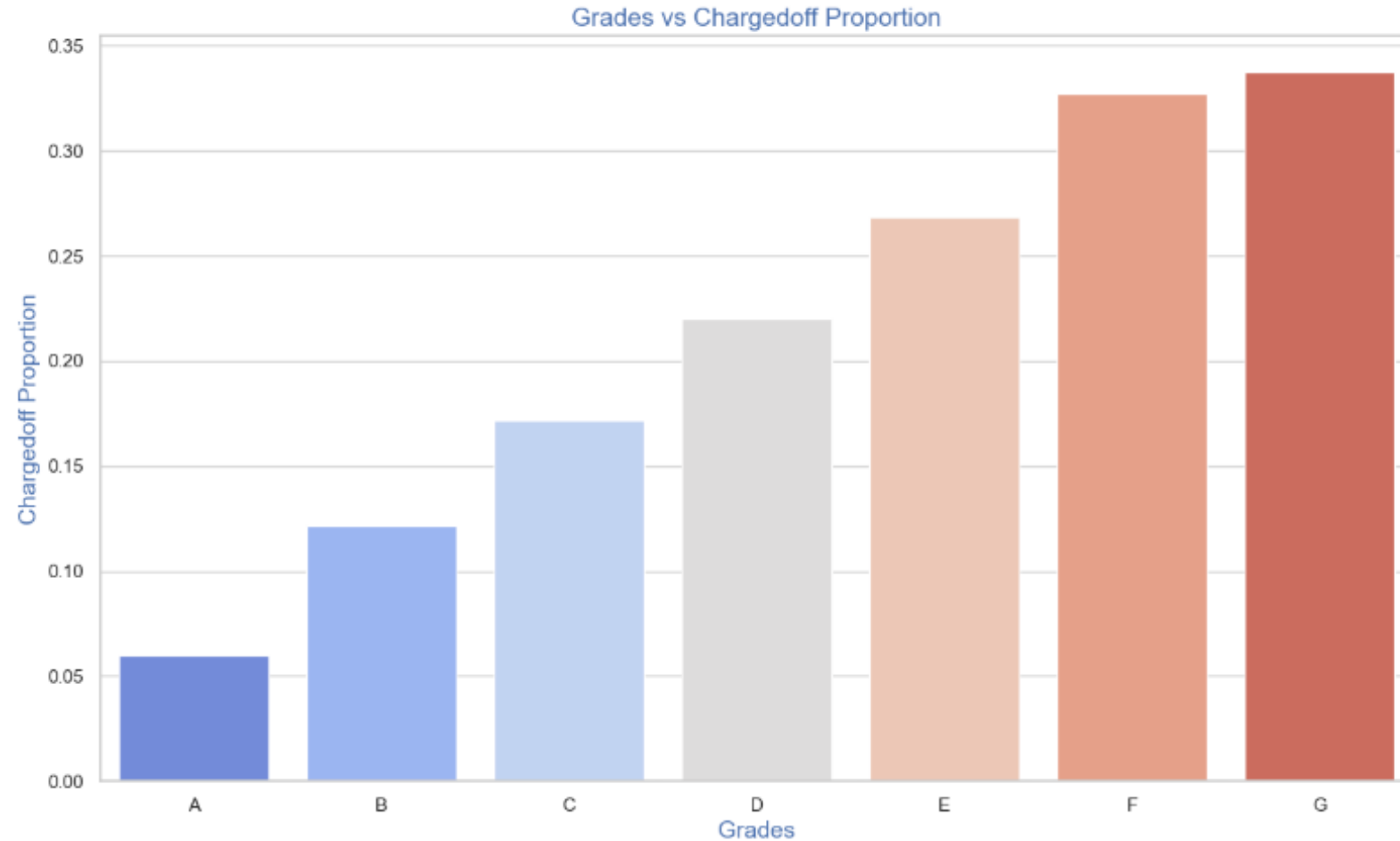
Observations: Grade "A" has very less chances of charged off. Grade "F" and "G" have very high chances of charged off. Chances of charged off is increasing with grade moving from "A" towards "G"

Bi-variate Analysis – Few important Columns

Grades vs Annual income
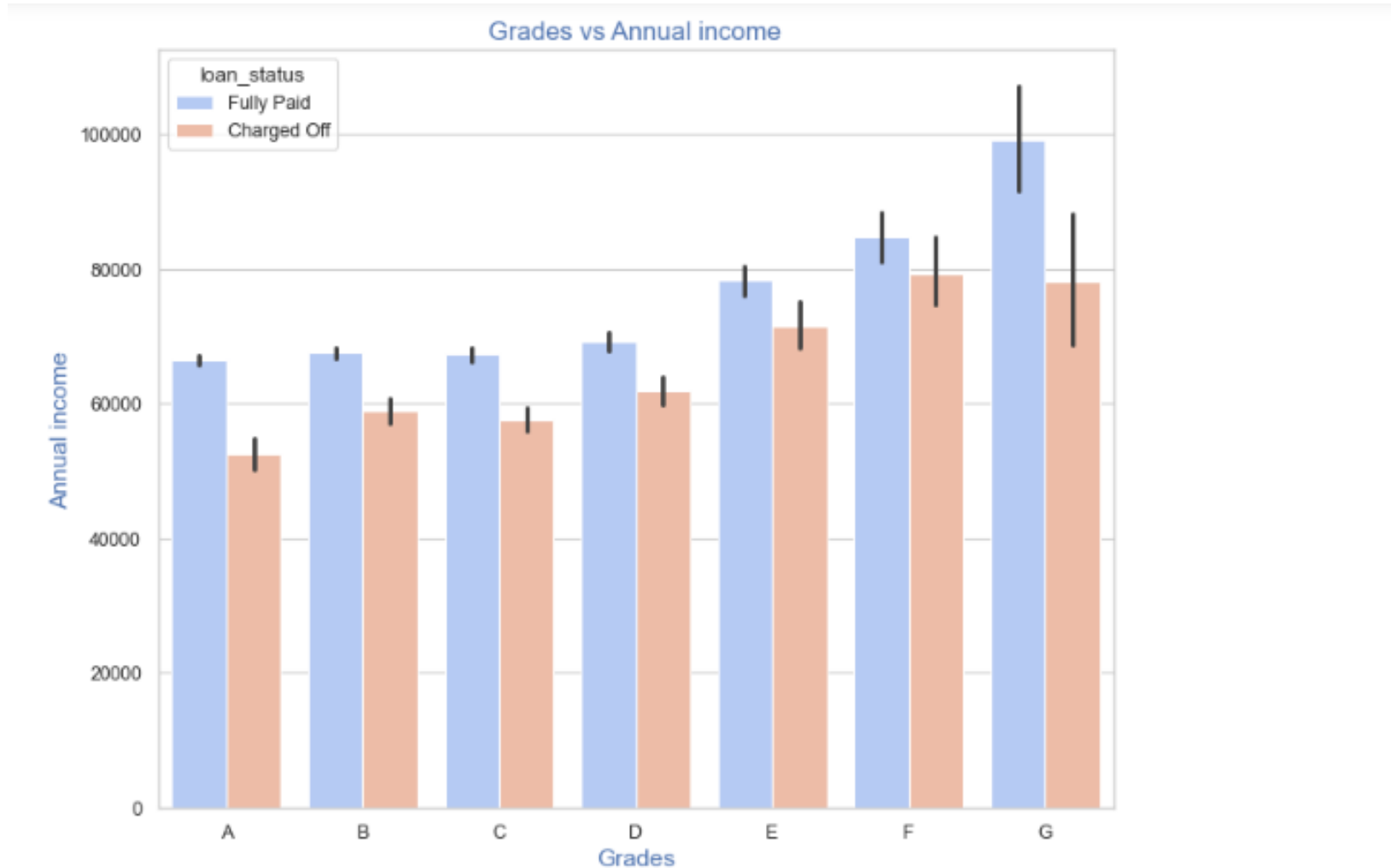
Observations: 'Charged off' is happening at Grade F than compared to other Grades

Bi-variate Analysis – Few important Columns

Employment Length vs Chargedoff Proportion
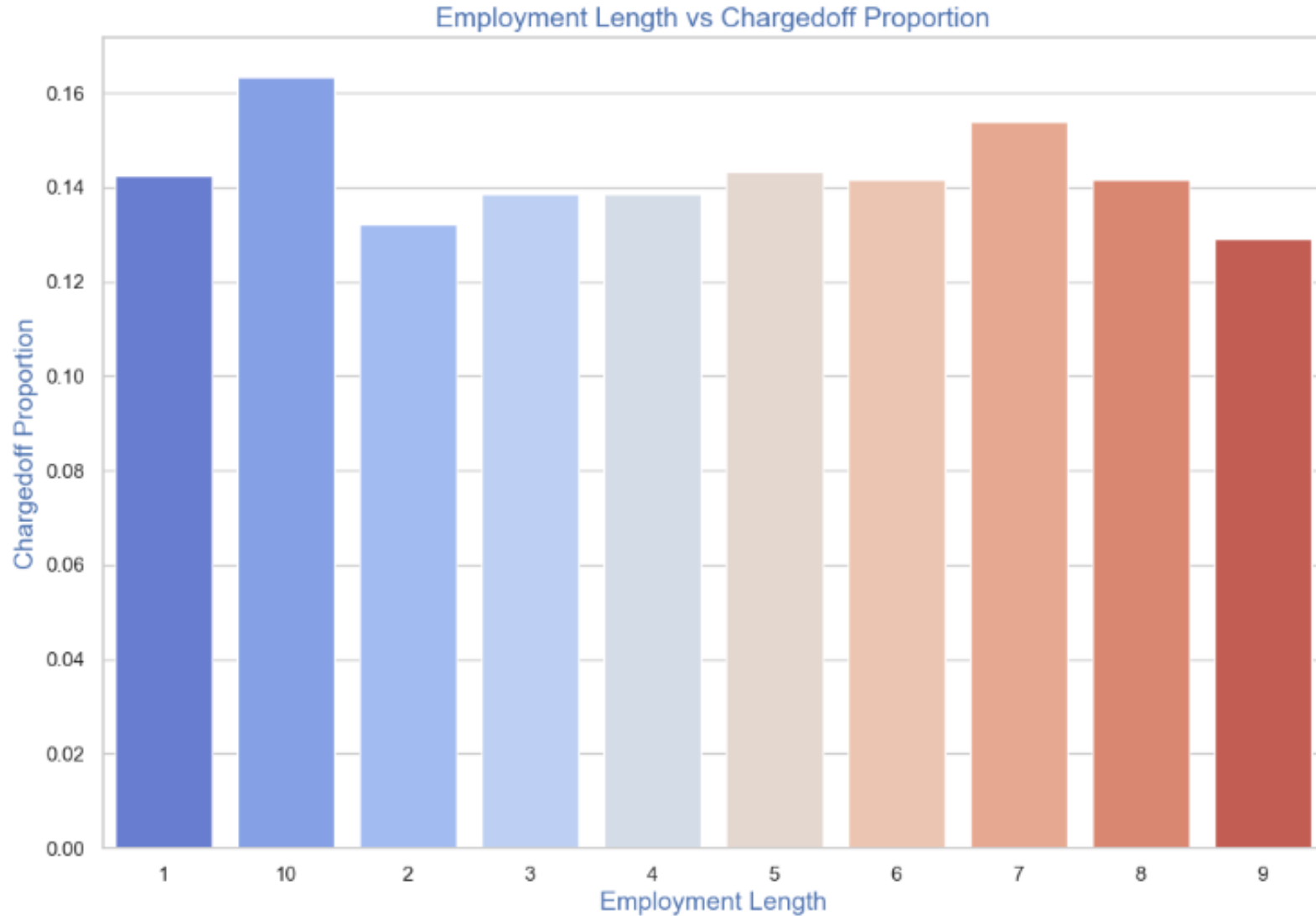
Observations: < 1 year of work experience have high chances of getting charged off since their income will be less to repay loan.

Bi-variate Analysis – Few important Columns

Loan Payment Term vs Interest Rate
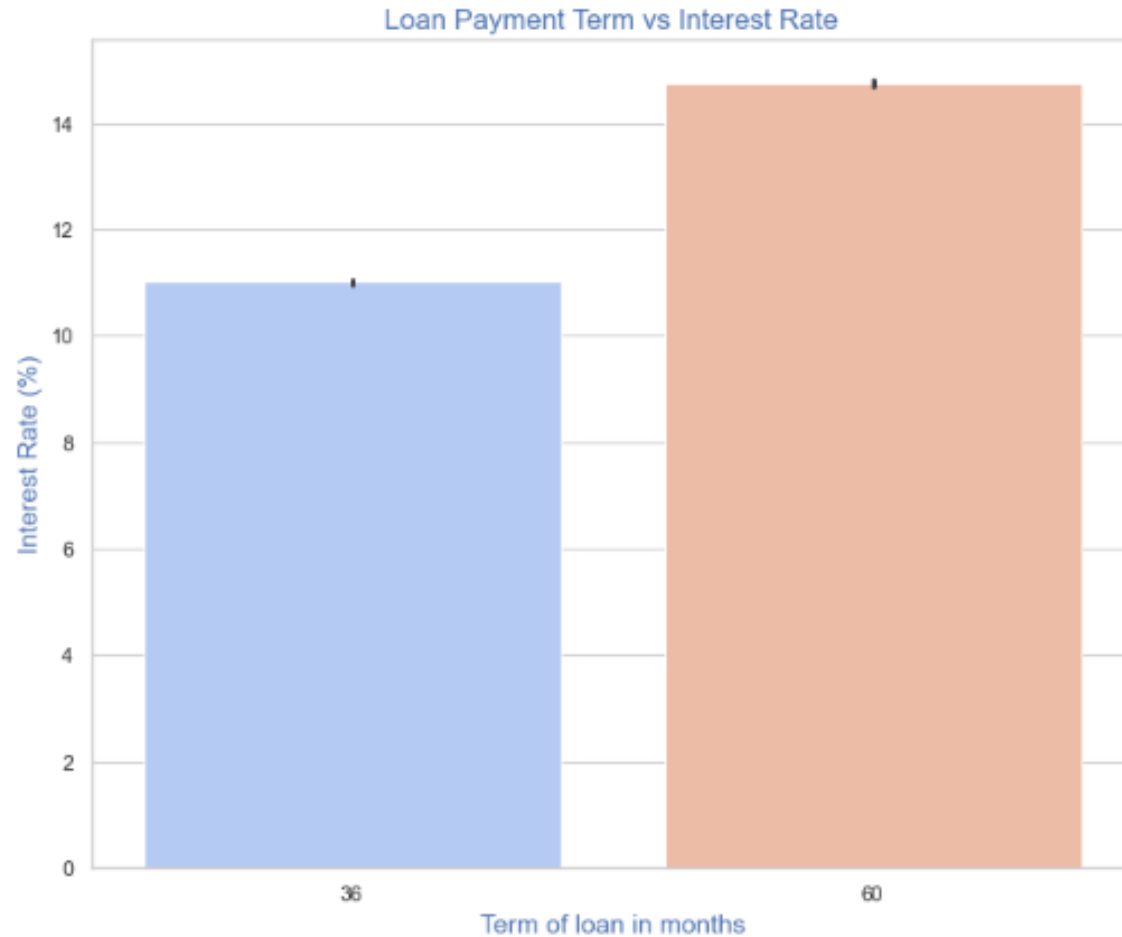
Observations: Interest rate is higher for 60 months loan term. Loans issued for longer term had higher interest rates for repayement

Bi-variate Analysis – Few important Columns

dti vs Interest Rate
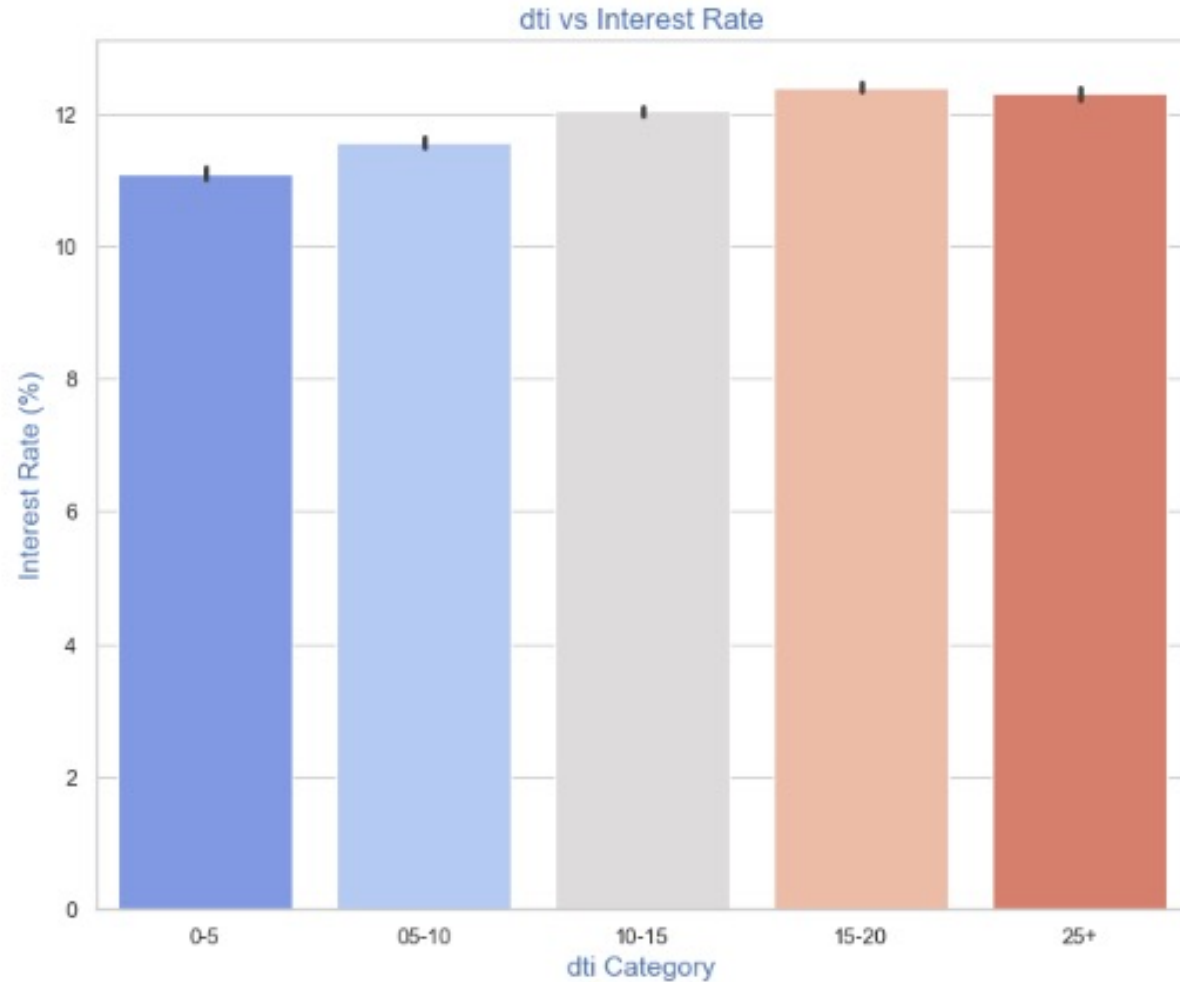
Observations: Borrower with Low "dti" may get a lower Interest rate. Slight Increase in Interest rate with increase in "dti".

Bi-variate Analysis – Few important Columns

Purpose of Loan vs Interest Rate
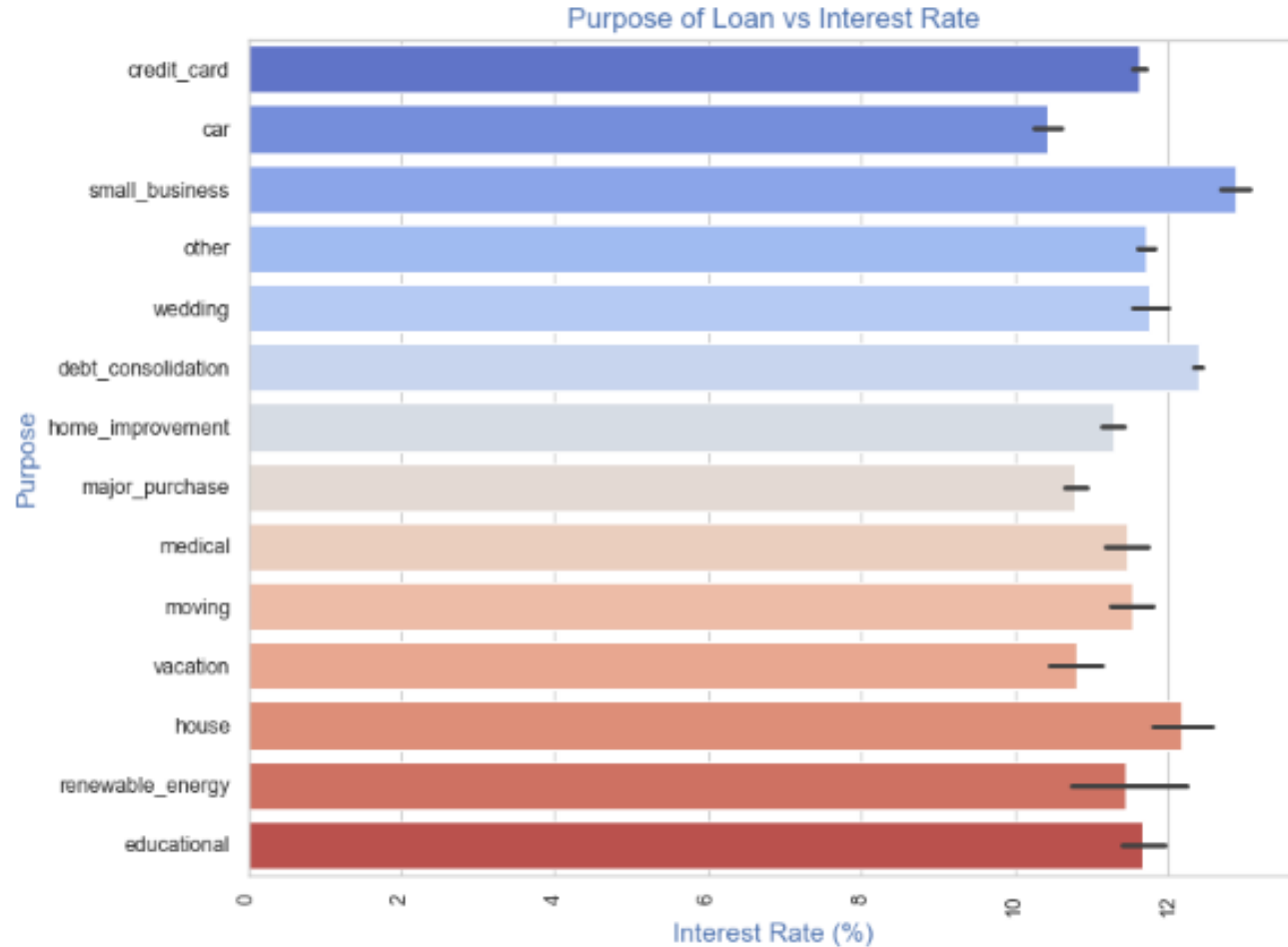
Observations: Interest rate is highest for small business purpose. Debt consolidation falls under 2nd where borrowers had to pay more interest rate.

Bi-variate Analysis – Few important Columns

Derogatory Public Record vs Chargedoff Proportion
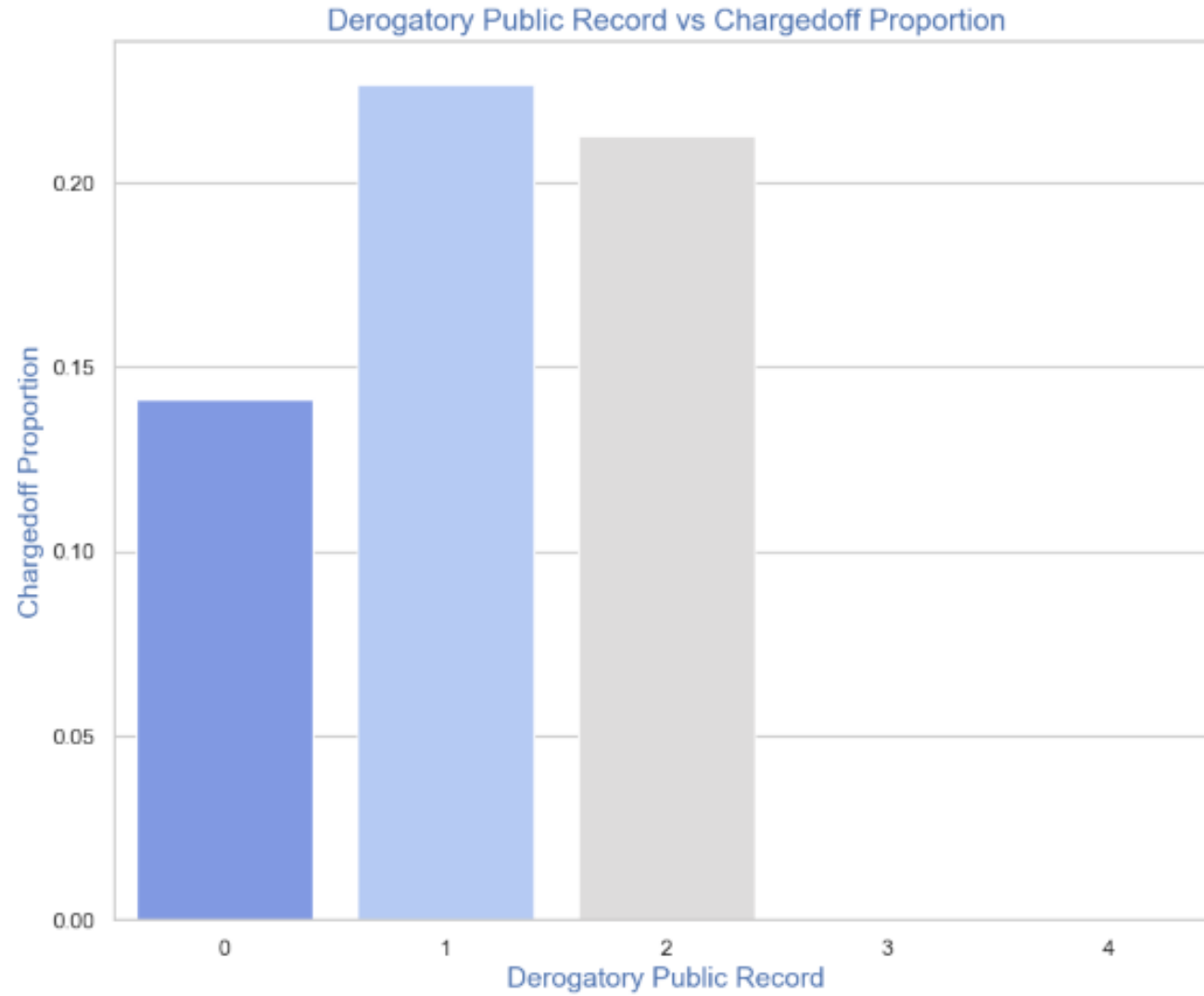
Observations: Pub_rec value 1 or 2 have charged off chances higher than who have no Derogatory Public Record. Pub_rec count 3-4 has less numbers so cannot reach on any conclusions.

Correlation Chart– Few important Columns

Correlation between numerical factors driving to Loan defaulters

Observations: Below mentioned details are having strong correlation

Increase in interest rate with loan amount

Increase in Loan defaulters with interest rate

Increase in loan amount provision with annual income

Final Summary & Suggestions:

Based on the Analysis done on the Variables, we conclude the below mentioned points :-

➢ Small Business Applicants have high chances of getting charged off.
➢ Charged off proportion increases with grades moving from "A" towards "G".
➢ Charged off proportion increases as Interest Rate Increases.
➢ Higher the public bankruptcy record greater the charged-off proportion.
➢ The loan amounts are bigger on average for small business purpose among all purposes of Loan.
➢ Those who already have Derogatory Public Records have higher charged off chances than others.
➢ Average interest rate is considerably higher for 60 months loan term than 36 months.
➢ Ones getting charged off have lower annual incomes than the ones who has fully paid for each and every grade.

Suggestions to Lending Club:

➢ Loans for Small Business Applicants should be checked properly.
➢ Loan approval should be avoided for those who already have Derogatory Public Records.
➢ Loan approval should be avoided for those who already have Public Bankruptcy Records.
➢ Loan approval for Low quality loans should be avoided or given for smaller loan repayment term.
➢ Lower annual income applicants should be avoided for big loan amounts with higher interest Rates.
➢ Loan approval should be avoided for applicants who doesn't have a source of income.