Lending Club Case Study

Submitted by - Manish Gautam & Gaurav Mehrotra

Problem Statement

You work for a consumer finance company that 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:

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.

The data given below contains information about past loan applicants and whether they 'defaulted' or not.

The aim is to identify patterns that indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of the loan, lending (to risky applicants) at a higher interest rate, etc.

Summary of the analysis

- Almost 14% of loans are charged off.
- Grade E, F, & G loans are the major defaulters and have been correctly graded. Grades are the strongest predictor of the default.
- Grade A has the least defaulters but still got some defaulters; grading could be done better to minimize the defaulters in Grade A to avoid business losses.
- Loans with subgrades of F and G are the major defaulters.
- Employment length has no impact on the defaulting, and we can see across the range of the experience that people are defaulting.
- A loan with the longer term has almost double the chance of defaulting compared to a short term; the company should promote short-term goals.
- Given numbers of loans were less in 2007, almost 29% got charged off.

Summary of the analysis

- February is the month where the least number of loans has been issued; this is in sync with the financial year cycle, because of March being the financial year end. This makes sense, and the last quarter of the year has the least number of loans issued.
- WY state has the least defaulters, and NY has the maximum; this could also be because NY is the financial capital, and more loans are issued in NY; hence, more defaulters. CA has the maximum loans but average defaulters.
- Home ownership is not a strong parameter to determine the defaulter.
- Small businesses have seen the maximum defaulters; hence, the company should do more due diligence before giving loans in this sector. Bigger purchases seem to have minimum defaulters; it could also mean people with higher incomes doing those purchases and paying up the loans in time.

Summary of the analysis

- Surprisingly, verified loans have defaulted more, which is serious, and the company could do better here and make the verification process work in favor of the company by avoiding bad loans; this could also indicate corruption in the verification process.
- People with higher annual incomes seem to be doing better, and companies should target loans to such profiles.
- The highest interest rate loans are defaulting more; this makes sense, as usually not-so-good credit score profiles get loans at higher interest rates and get defaulted. The company could avoid such loans and make the best of the loans where interest rates are lowest.
- Bigger loans seem to be defaulting more; the company can do more scrutiny here and try to bring down the loan amount for better outcomes for the company.
- People with the maximum income are going for the maximum loans, which is good for the company as they default less.

Grades & subgrades are the strongest predictor of the default.
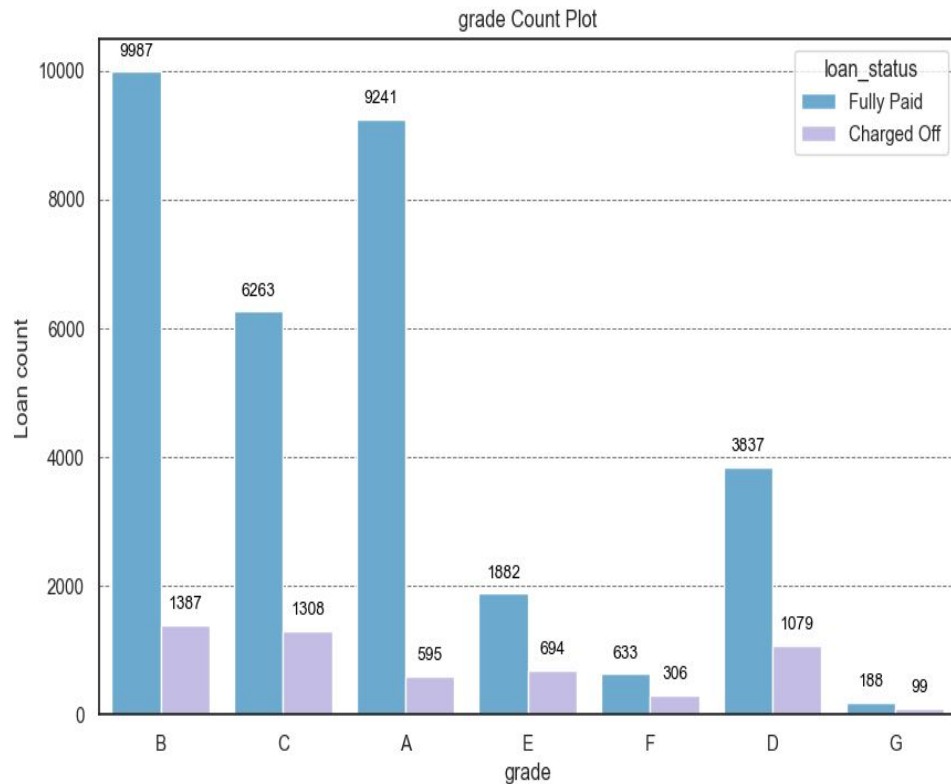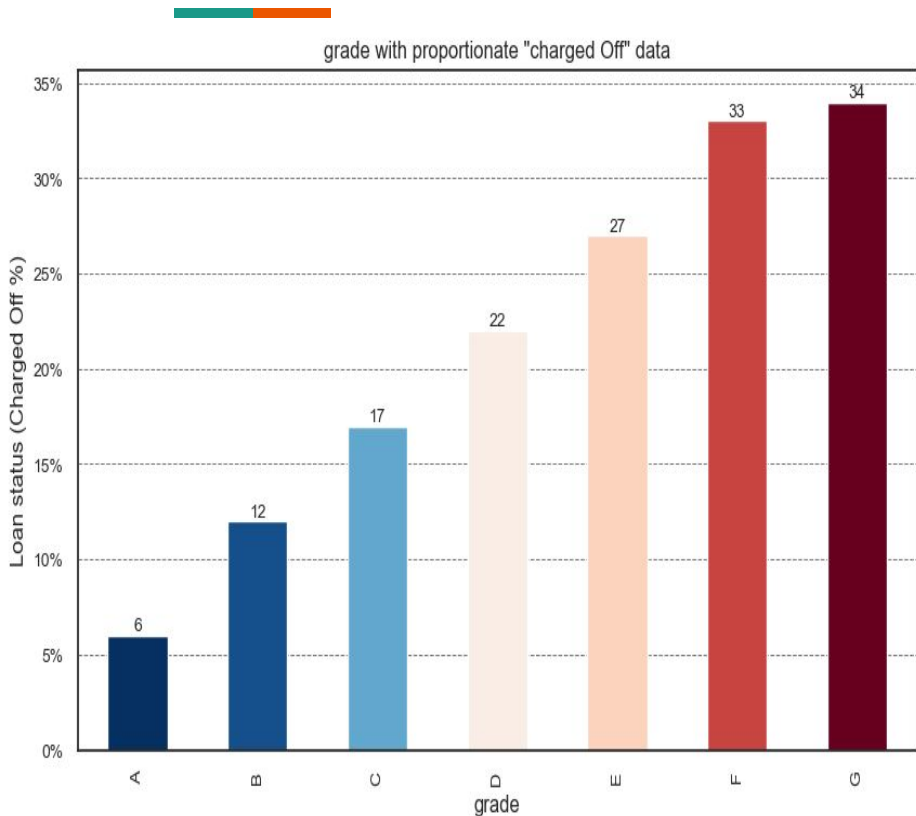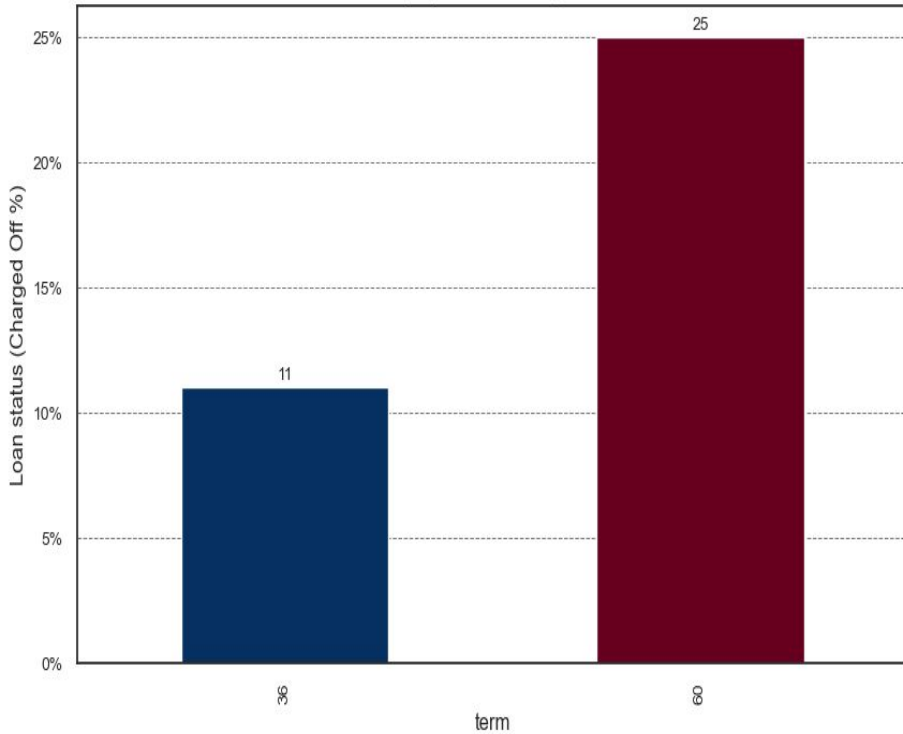
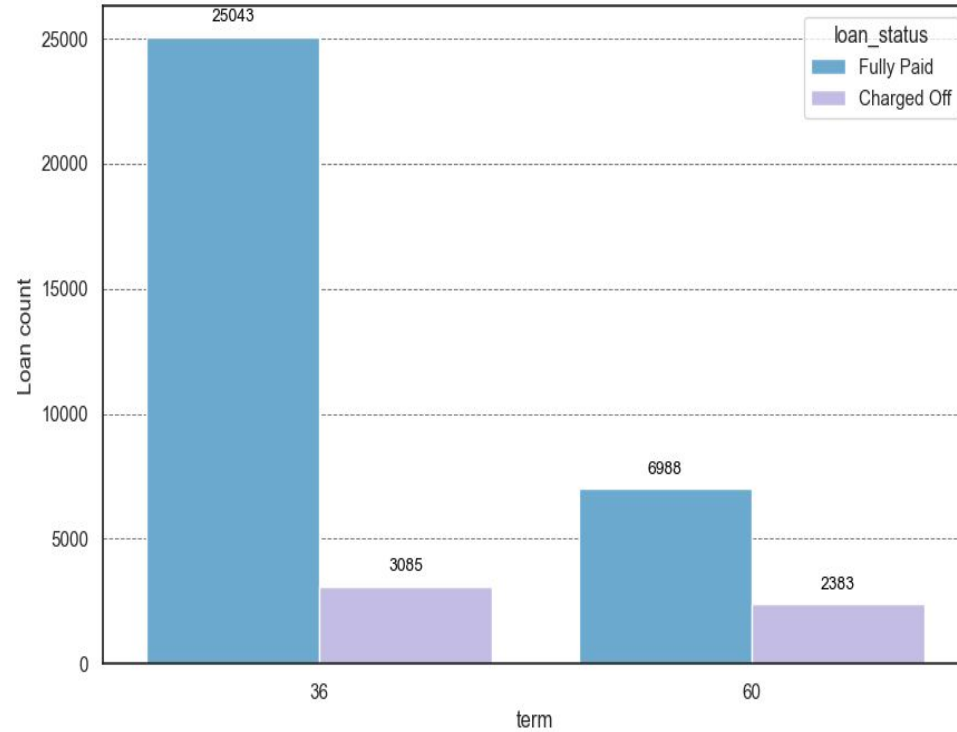# Grades & subgrades are the strongest predictor of the default.
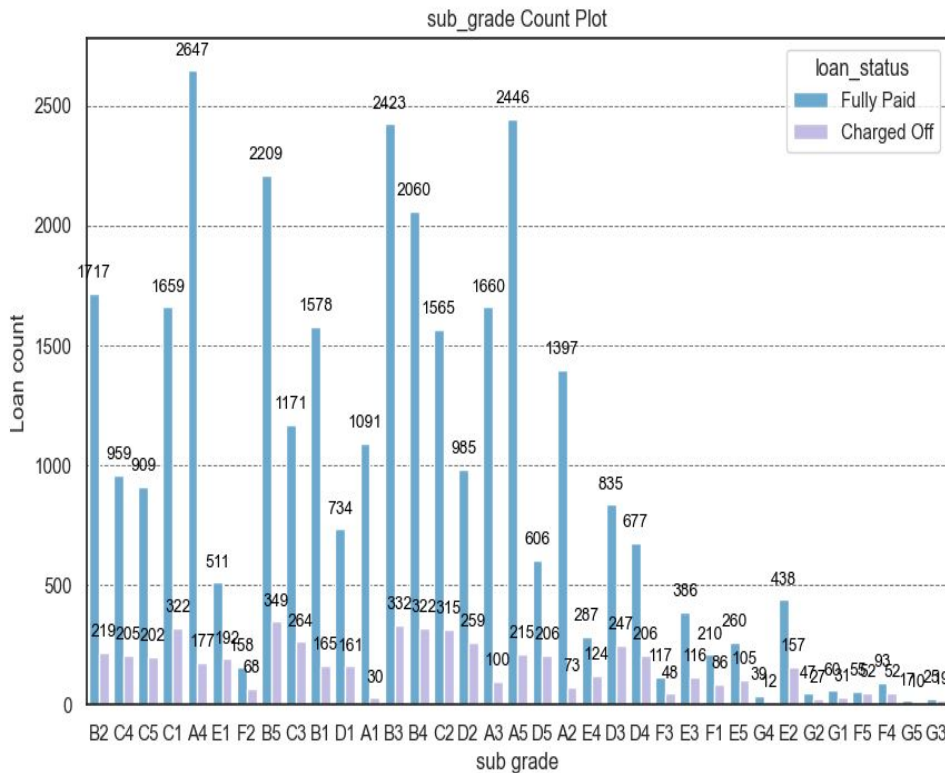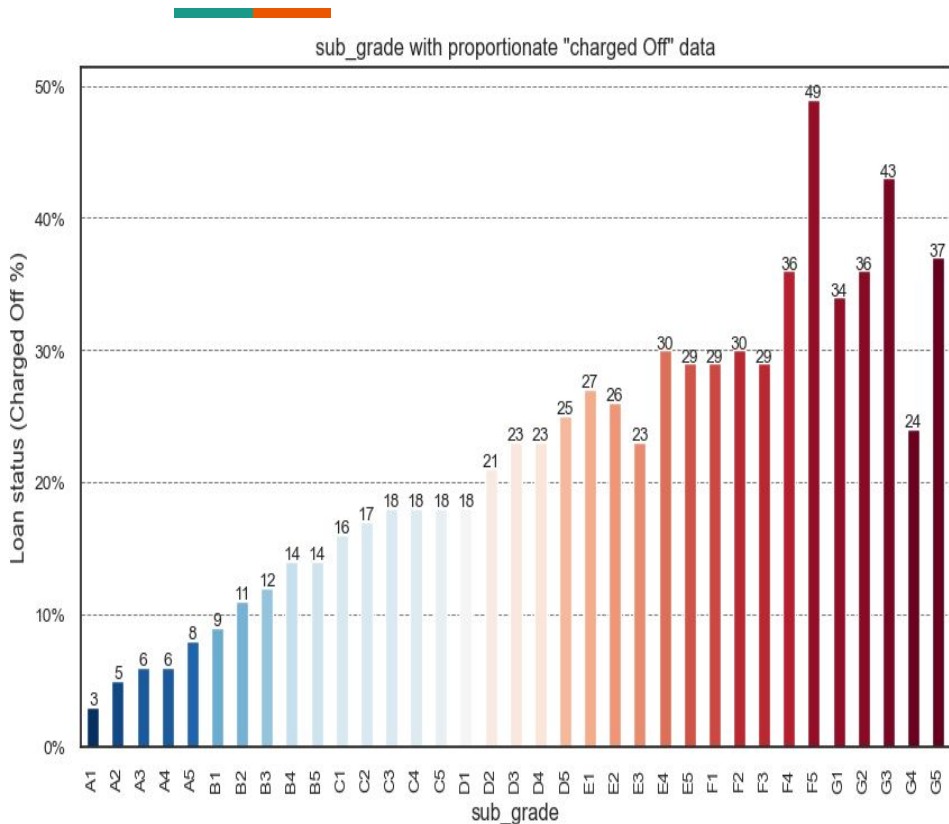
The longer- term loans have almost double the chance of defaulting compared to short-term ones; the company should promote short-term goals.
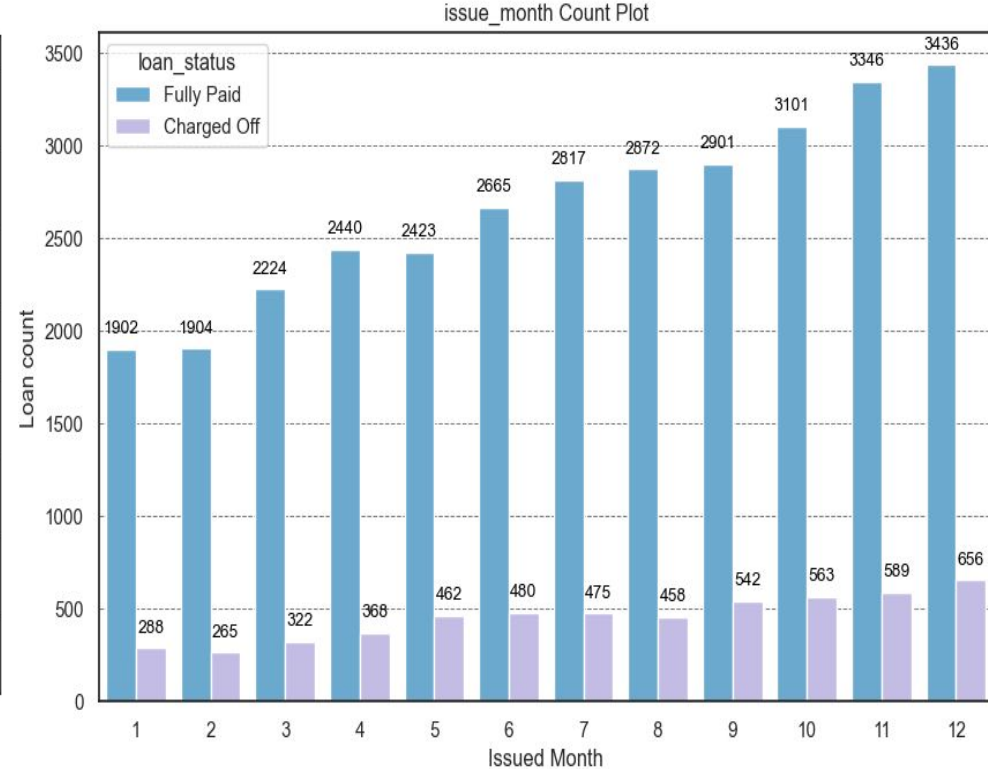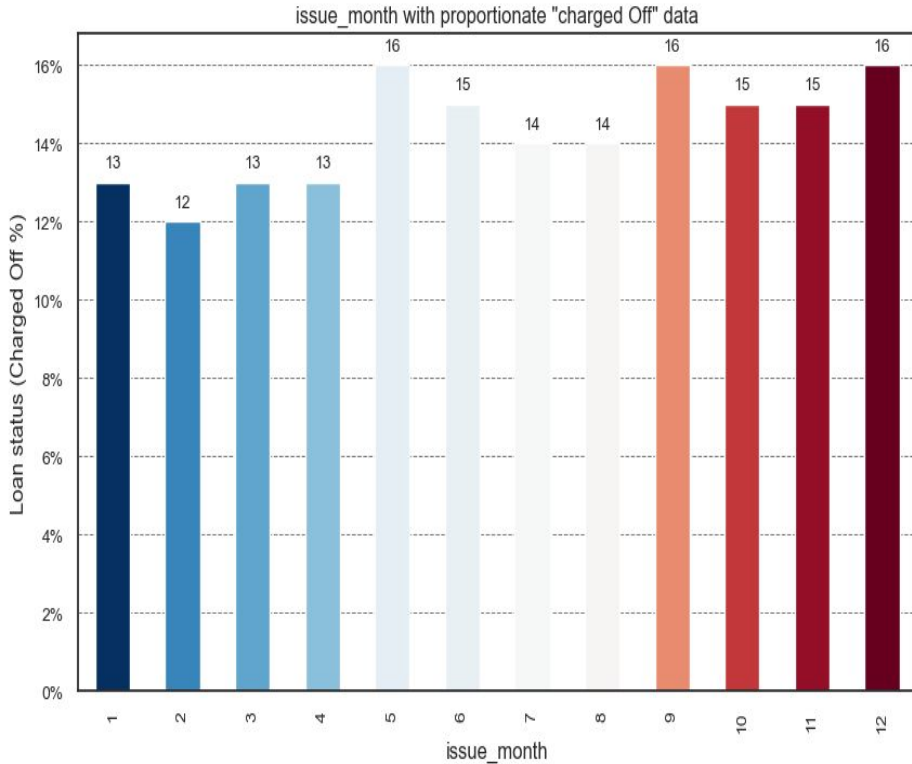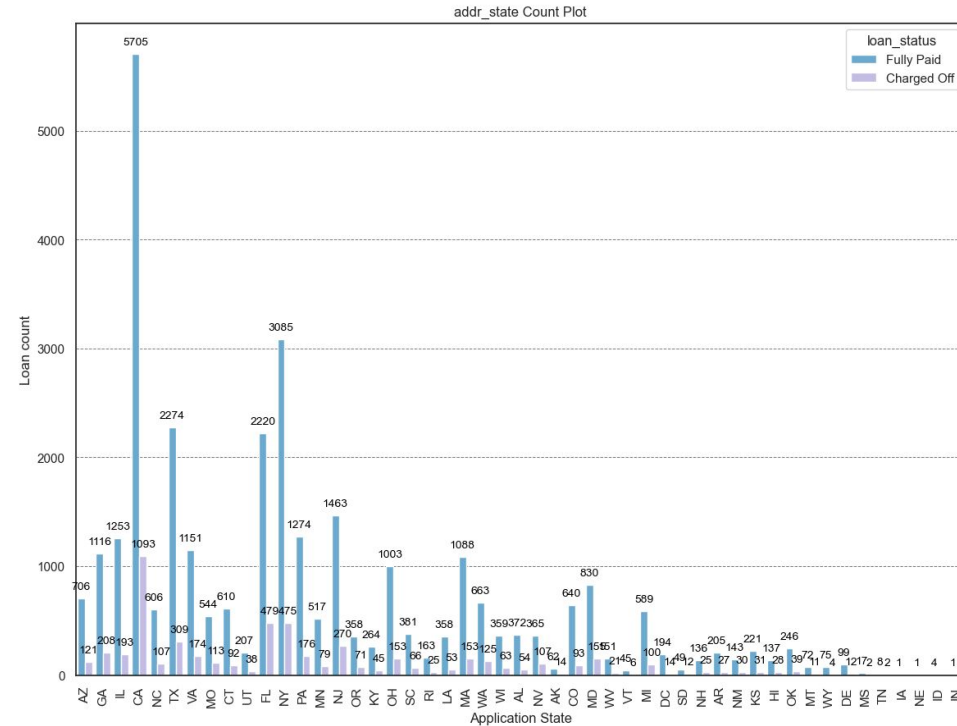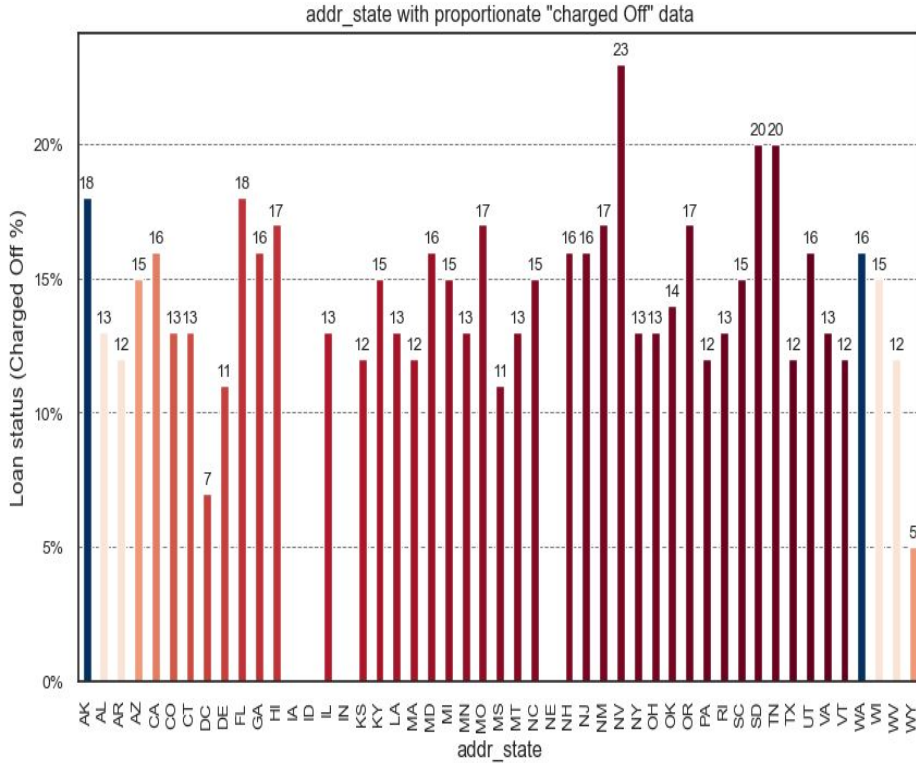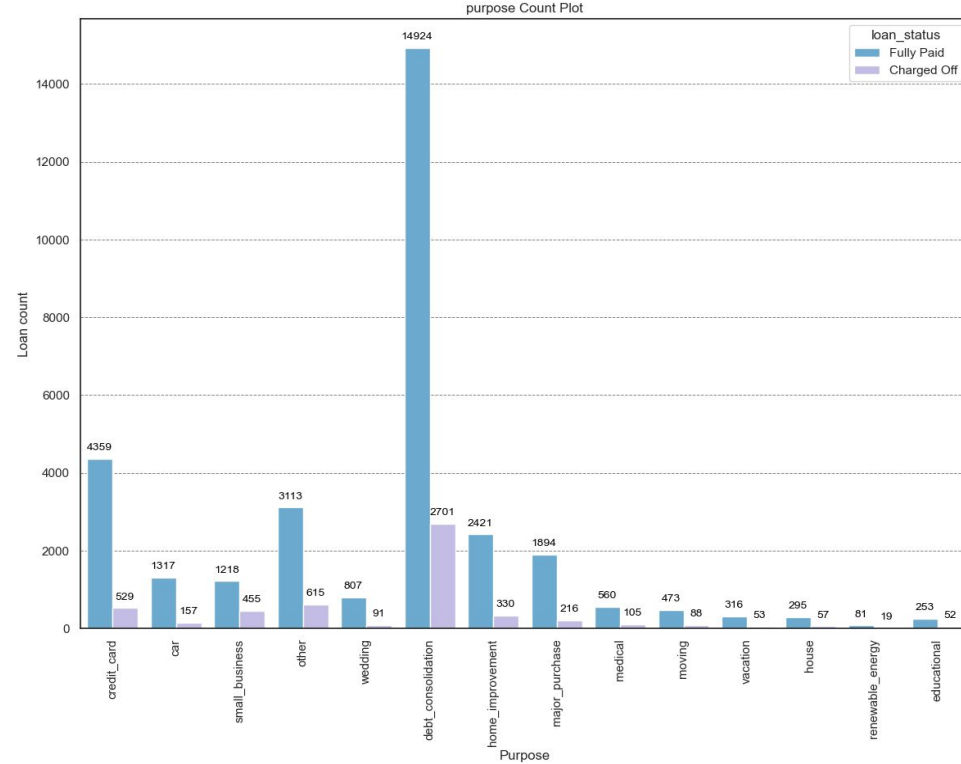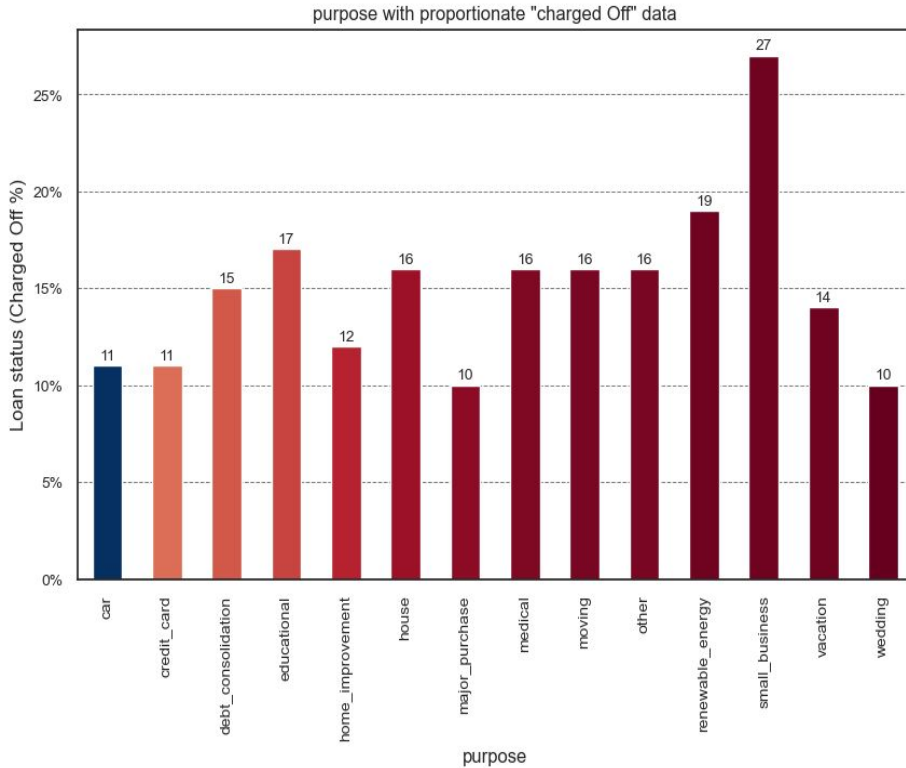
February and the last quarter of the year got the least number of loans issued; this is in sync with the financial year cycle.

issue_month with proportionate "charged Off" data
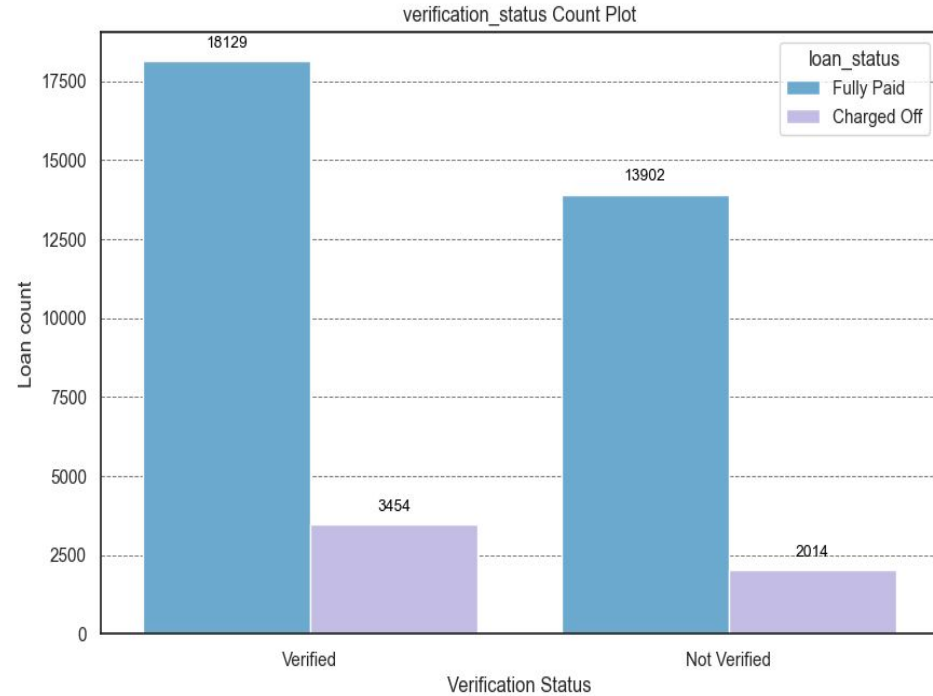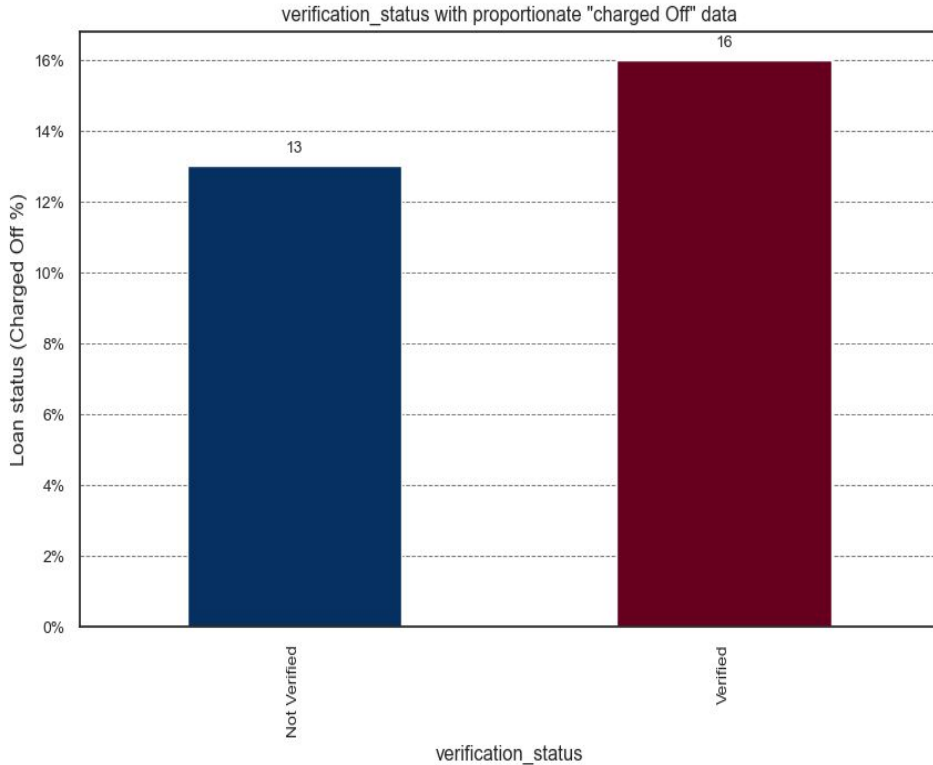

issue_month Count Plot

WY state has the least defaulters, and NY has the maximum; NY is the financial capital and hence has more defaulters. CA has the maximum loans but average defaulters.
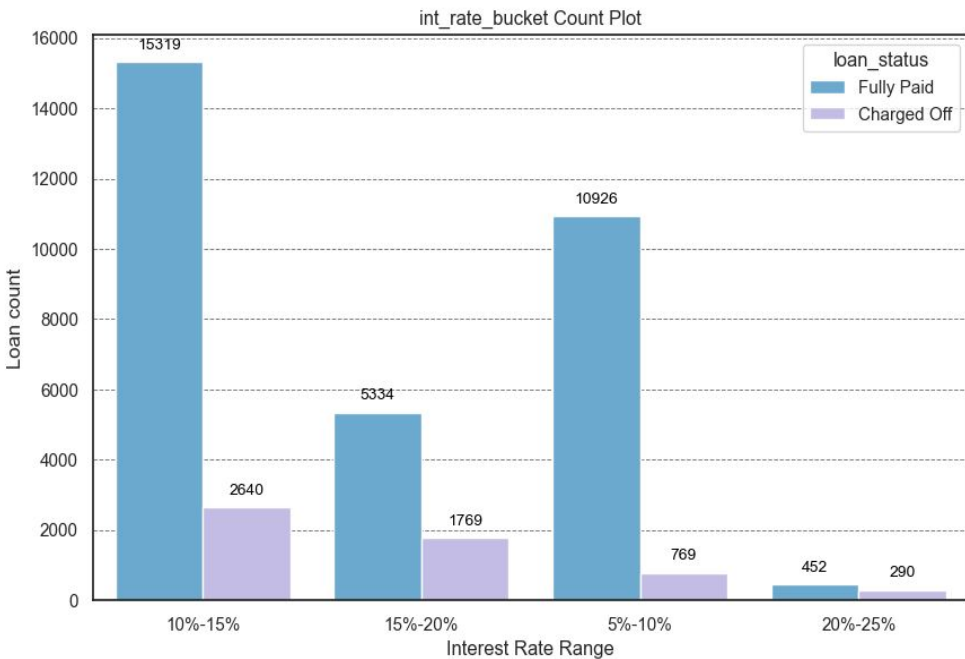
Small businesses have seen the maximum defaulters; hence, the company should do more due diligence before giving loans in this sector.

purpose with proportionate "charged Off" data

purpose Count Plot

Surprisingly, verified loans have defaulted more, which is serious, and the company could do better here and make the verification process work in favor of the company by avoiding bad loans.

verification_status with proportionate "charged Off" data

verification_status Count Plot

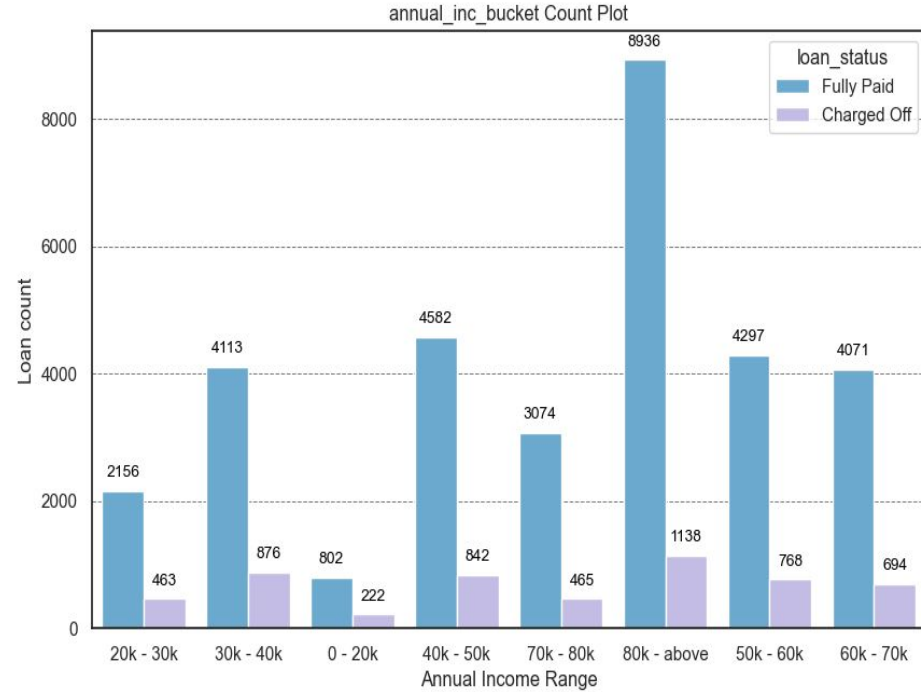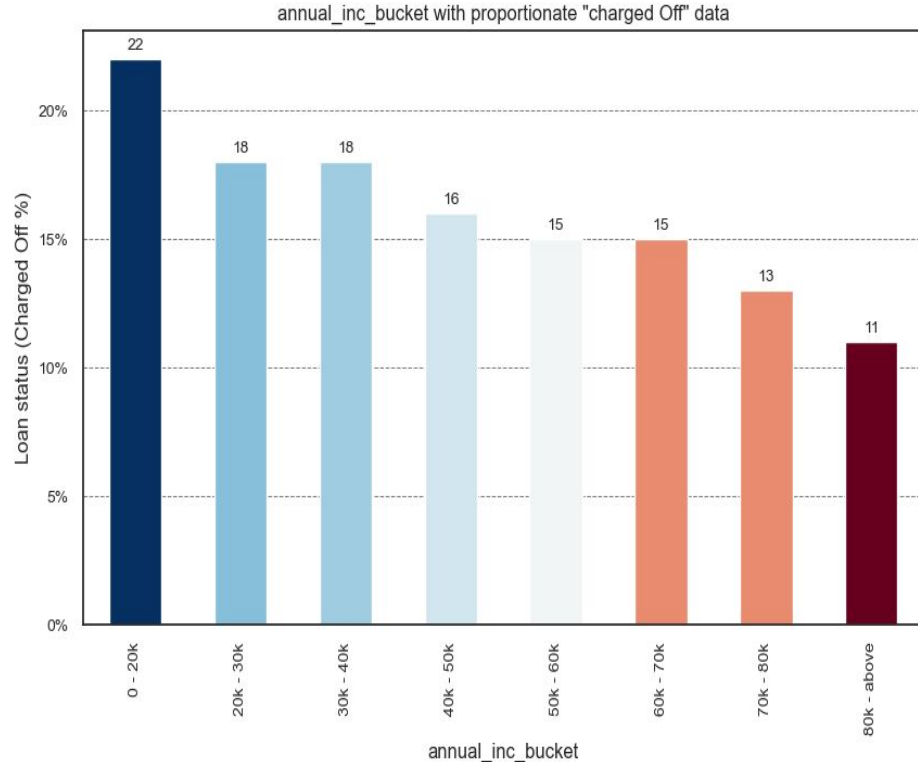**The highest interest rate loans are defaulting more; the company could avoid such loans and make the best of the loans where interest rates are lowest.**

Bigger loans seem to be defaulting more; the company can do more scrutiny here and try to bring down the loan amount for better outcomes for the company.

loan_amnt_bucket with proportionate "charged Off" data

loan_amnt_bucket Count Plot

People with the maximum income are going for the maximum loans, which is good for the company as they default less.

annual_inc_bucket with proportionate "charged Off" data

Loan status (Charged Off %) — annual_inc_bucket

Bars: 0 - 20k: 22, 20k - 30k: 18, 30k - 40k: 18, 40k - 50k: 16, 50k - 60k: 15, 60k - 70k: 15, 70k - 80k: 13, 80k - above: 11

annual_inc_bucket Count Plot

loan_status: Fully Paid, Charged Off

Loan count — Annual Income Range

20k - 30k: 2156 / 463, 30k - 40k: 4113 / 876, 0 - 20k: 802 / 222, 40k - 50k: 4582 / 842, 70k - 80k: 3074 / 465, 80k - above: 8936 / 1138, 50k - 60k: 4297 / 768, 60k - 70k: 4071 / 694

Correlation Metrics

Correlation Metrics Insights

- Installment has a strong correlation with loan_amnt, funded_amnt, and annual income.
- term` has a good correlation with interest rate

- Employment length has a very weak correlation with most of the parameters.
- Issues month and year also have no correlation with any parameter.