Lending Club Case Study

Submission

By

Ashwani Kumar

&

Anirudhya Bhattacharya

The problem

Company

Lending Club is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures.

Borrowers can easily access lower interest rate loans through a fast online interface.

Context

Lending Club wants to understand the **driving factors** behind loan default, i.e. the **driver variables** which are strong indicators of default.

The company can utilise this knowledge for its portfolio and risk assessment.

Problem statement

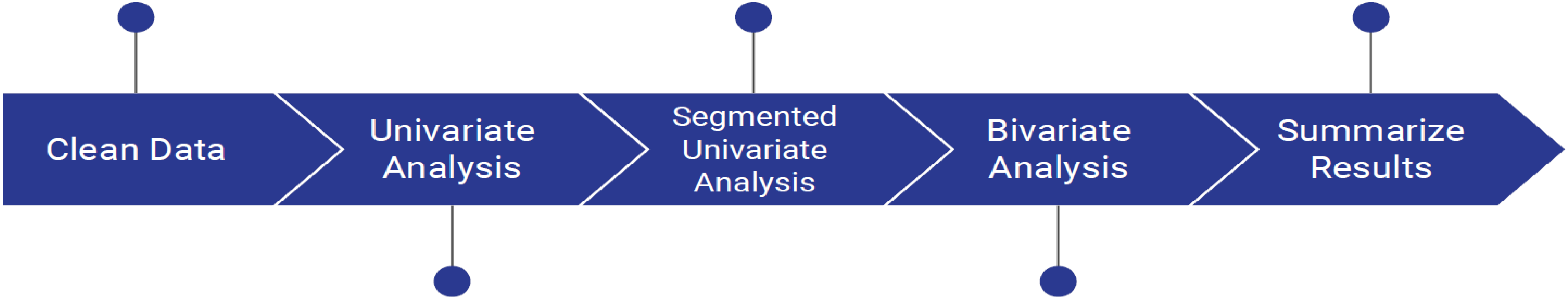
As a data scientist working for Lending Club analyze the dataset containing information about past loan applicants using EDA to understand how *consumer attributes* and *loan attributes* influence the tendency of default

Analysis Approach

> Drop columns with null values, all random values or single category value
> Convert values to proper int, float, date representations

> Analyze variables against segments of other variables
> Create derived variables

Publish insights and observations



> Check distributions and frequencies of various numerical and categorical variables
> Create derived variables

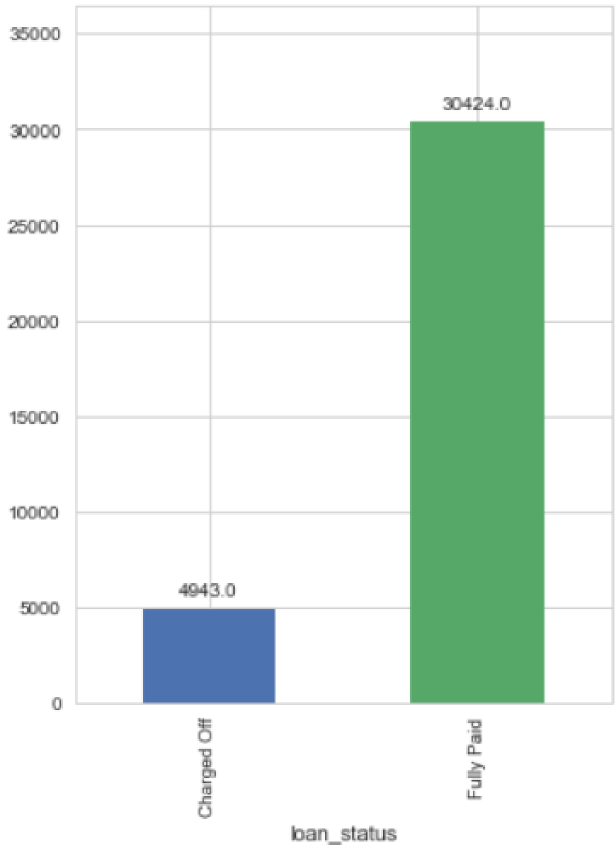
> Do correlation analysis
Check how two variables affect each other or a third variable
> Analyze joint distributions

Analysis - Overall Loan Status

Total Loans

Approximately 14% of loans are defaulted

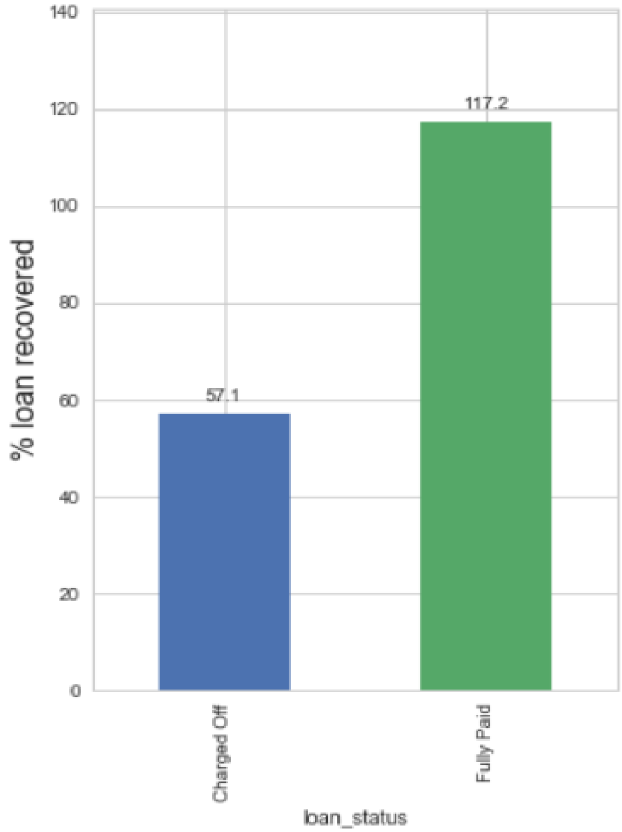
Any variable that increases percentage of default to higher than 16.5% should be considered a business risk.



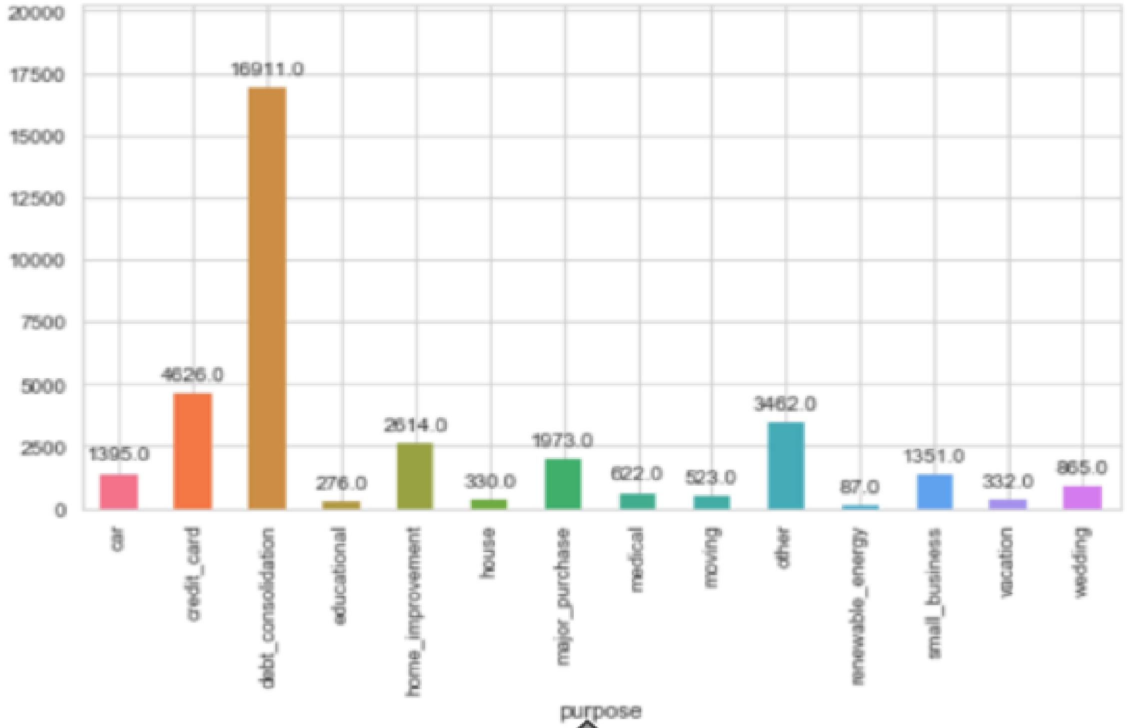
Total Money Earned

Lending Club only recovers 57% of the loan amount when loans are **defaulted**.

On fully paid up loans, the company makes 17% profit.

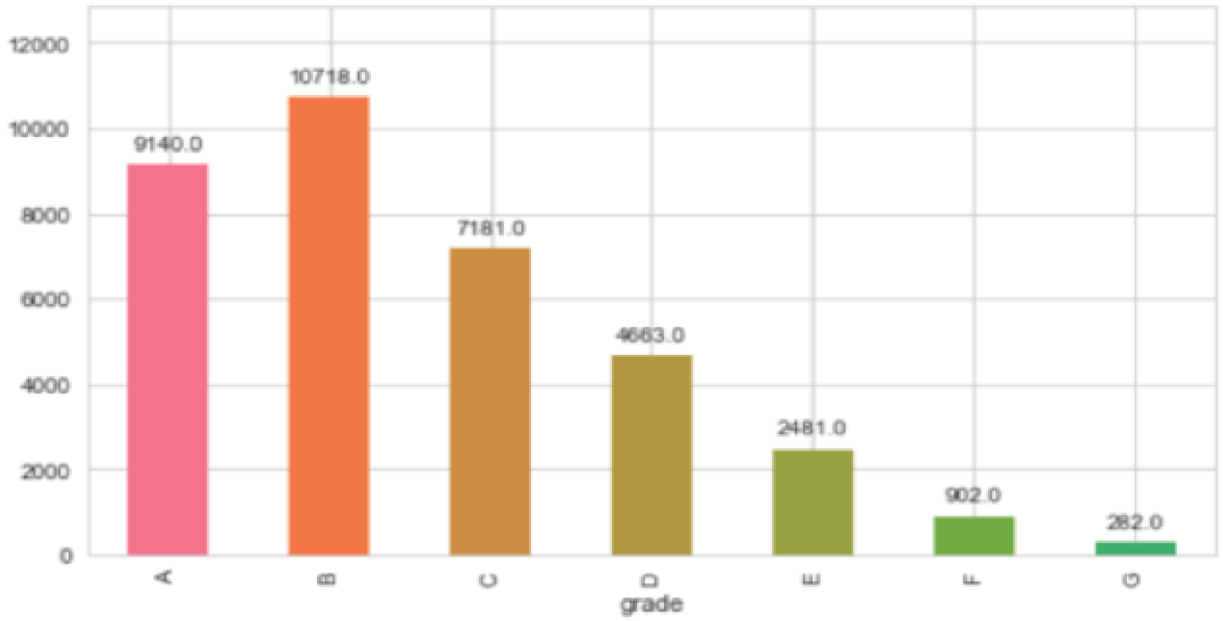


Analysis - Understanding Loans



Maximum number of loans are for debt consolidation, followed by credit card

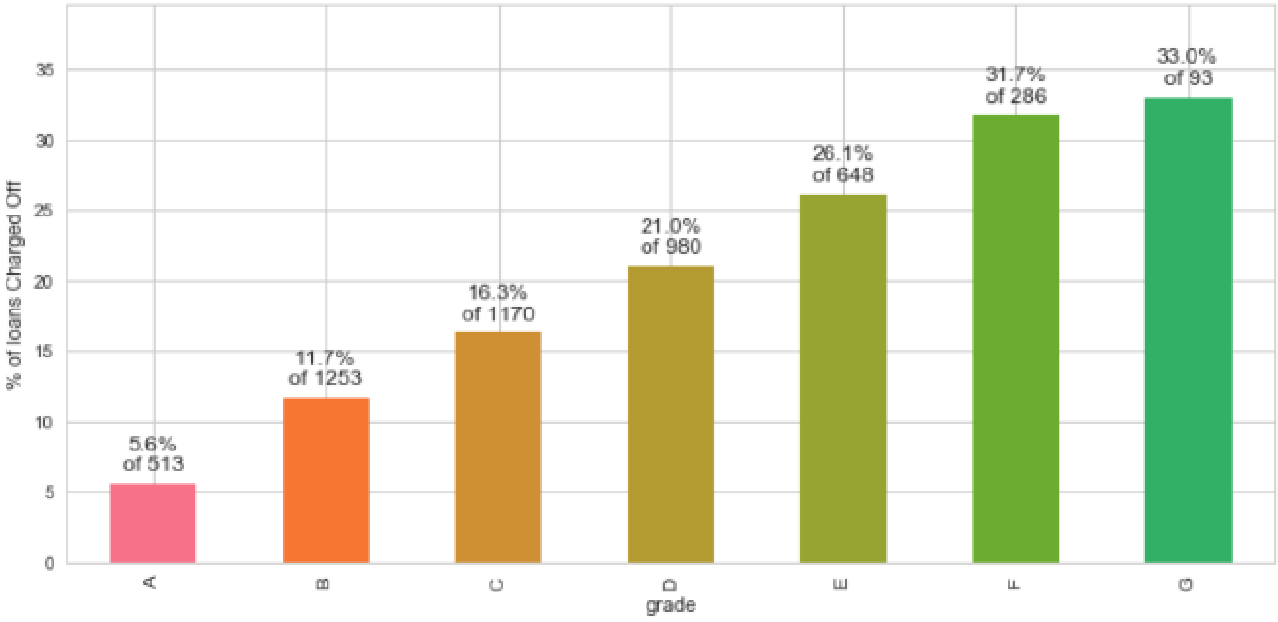
Most loans are high quality, with a grade of A or B

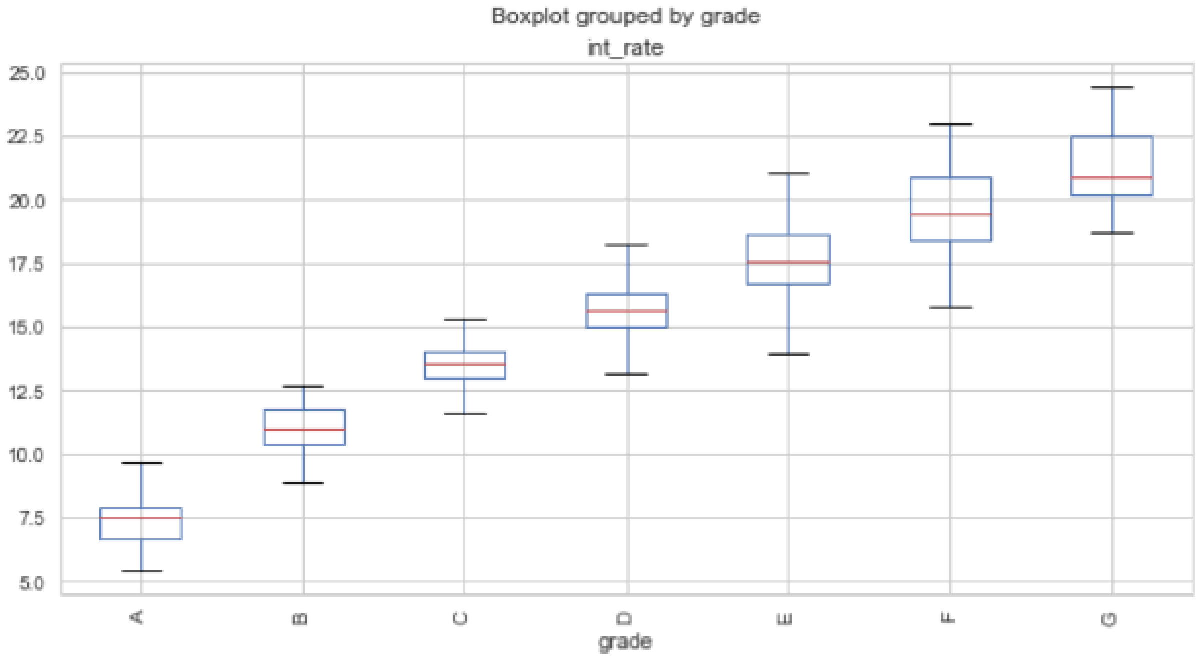


Analysis - Understanding Loans Continued

Lower grades have higher incidence of defaults on loans.
The grading system is working!

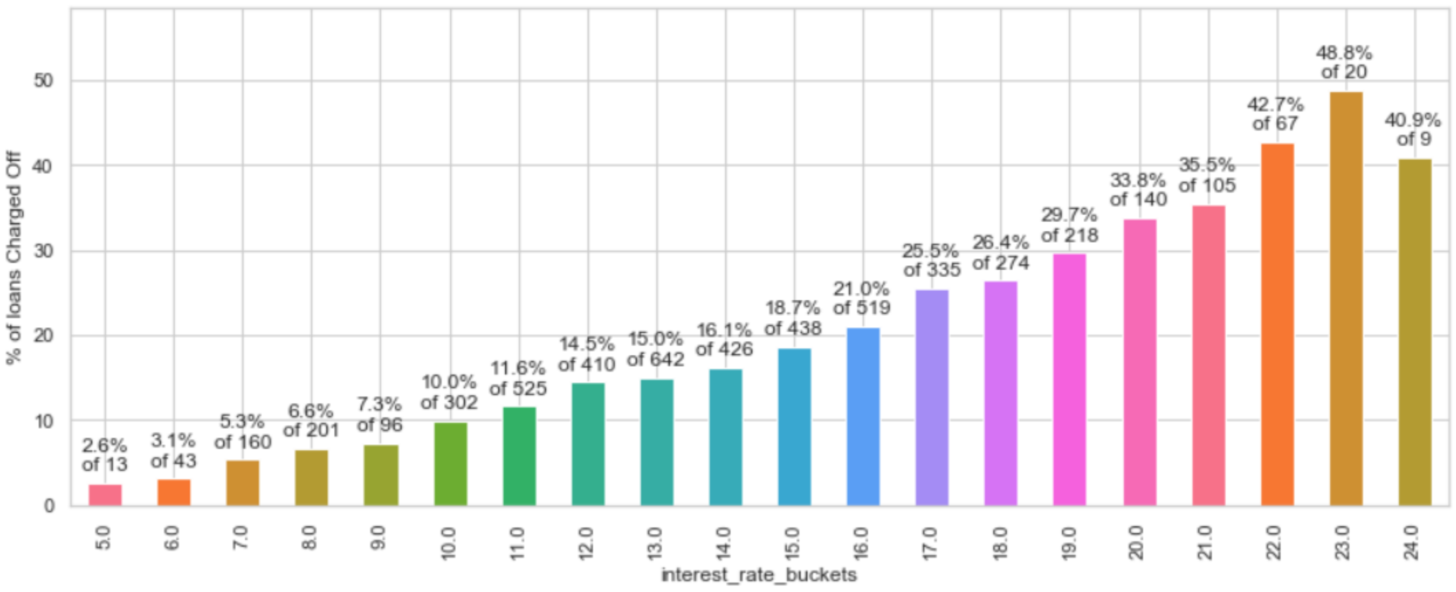
Lending Club charges higher interest rates as the grade of loan becomes worse.
However, as we will see on next slide - the driving variable for defaults is the higher interest rate.

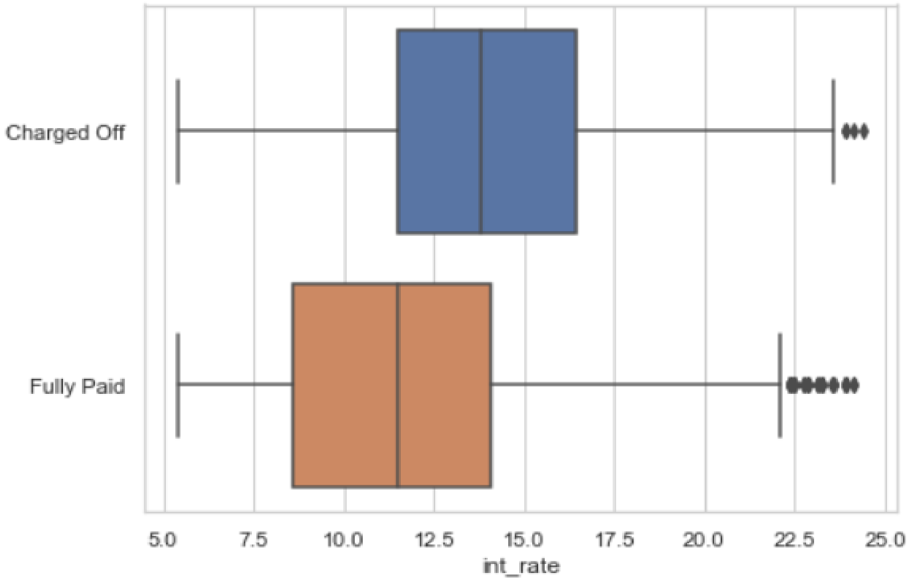




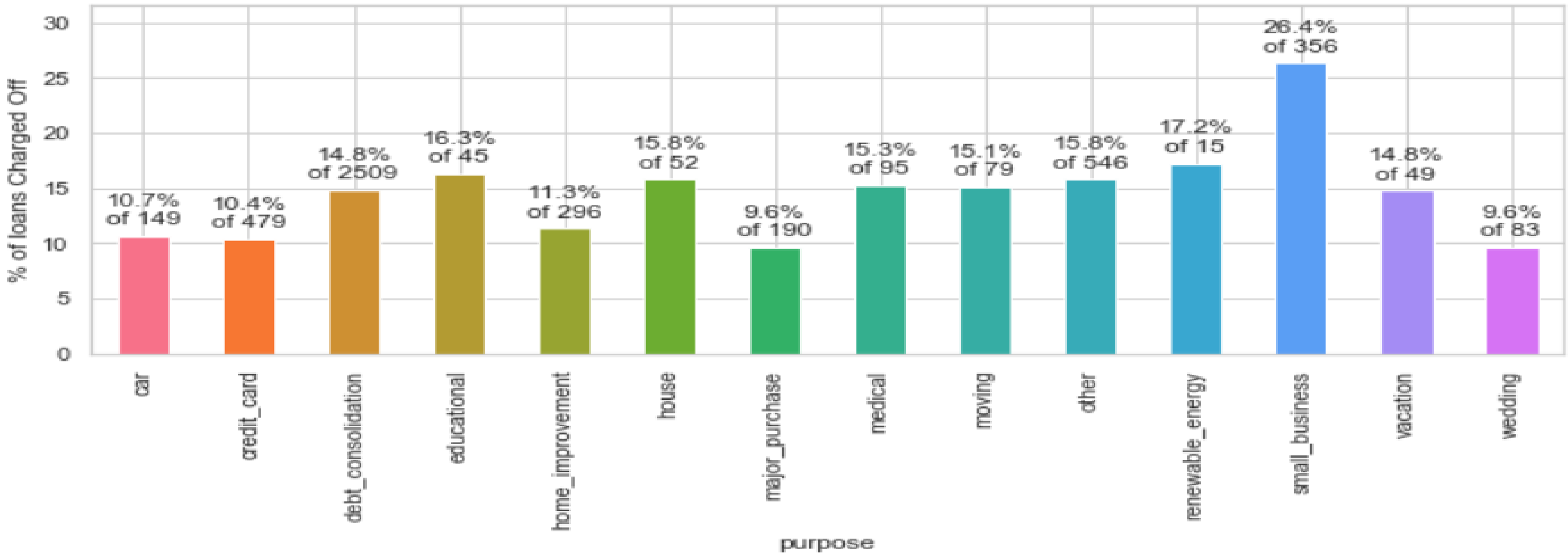
Analysis - Defaults by Interest Rate

Percentage of Defaults increases monotonically with higher interest rates. At rates of 19% and above, more than 30% of loans are Charged Off.





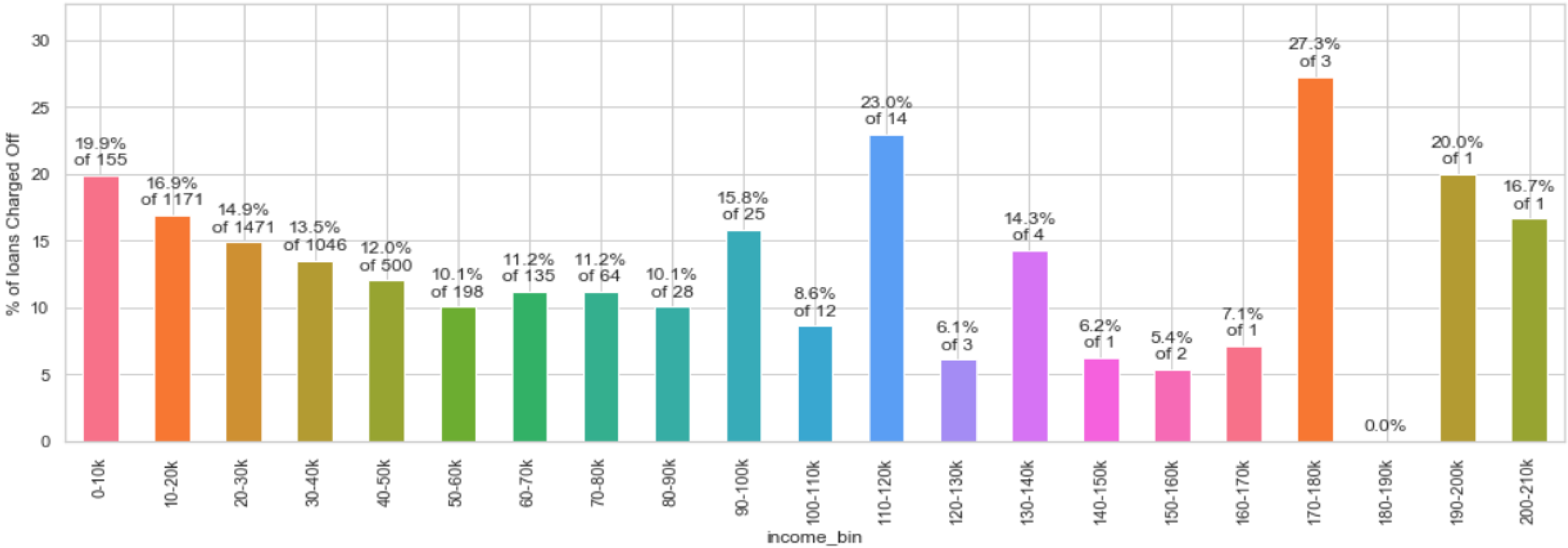
Analysis - Defaults by Loan Purpose



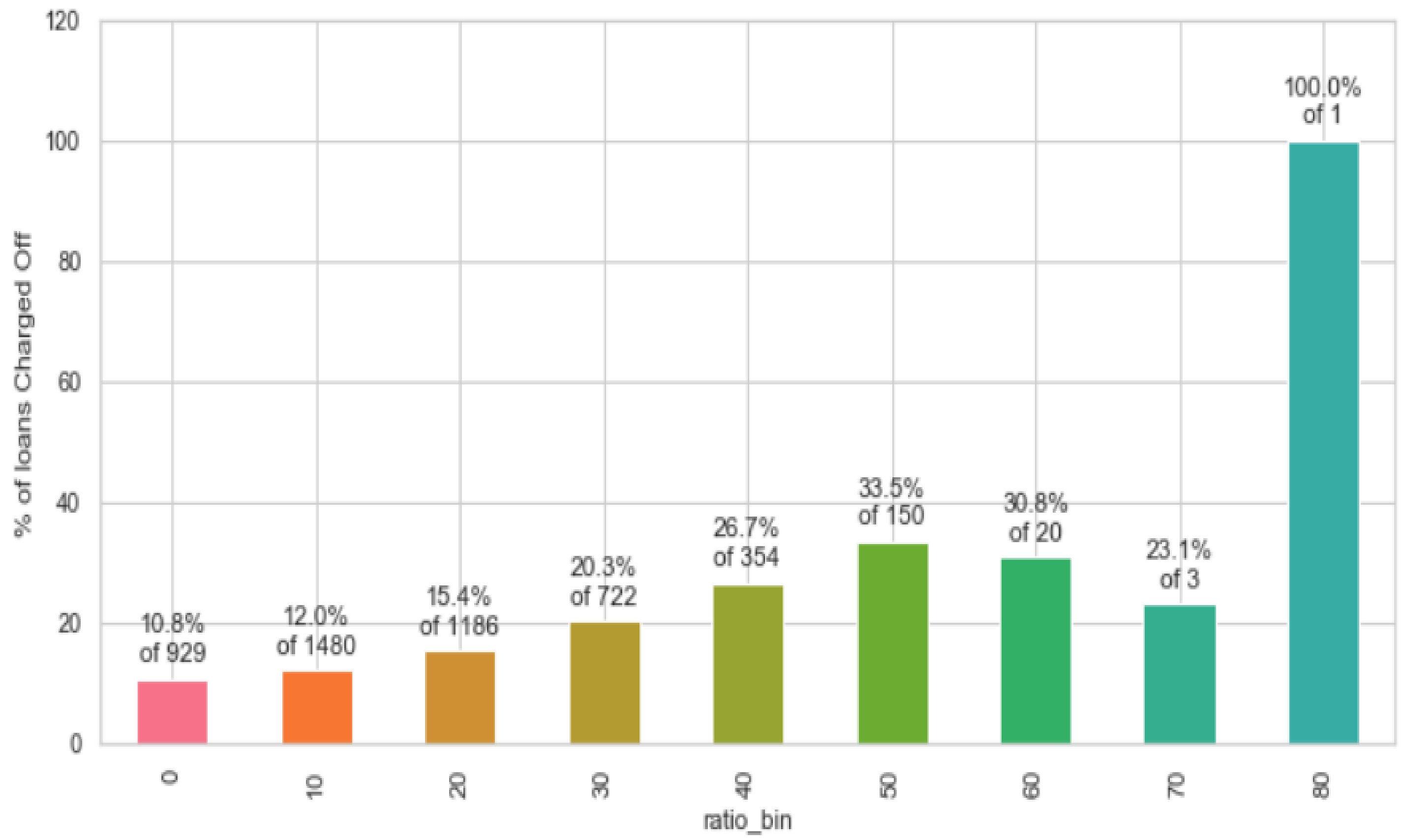
More than a quarter of loans taken for the purpose of running a small business are Charged Off.

Analysis - Defaults by Borrower's Income

- Borrowers having annual income less than 20000 default on their loans at much higher rates.
- Loan default decreases with higher annual income.
- As we will see on next slide – the ratio of amount to income is more important.

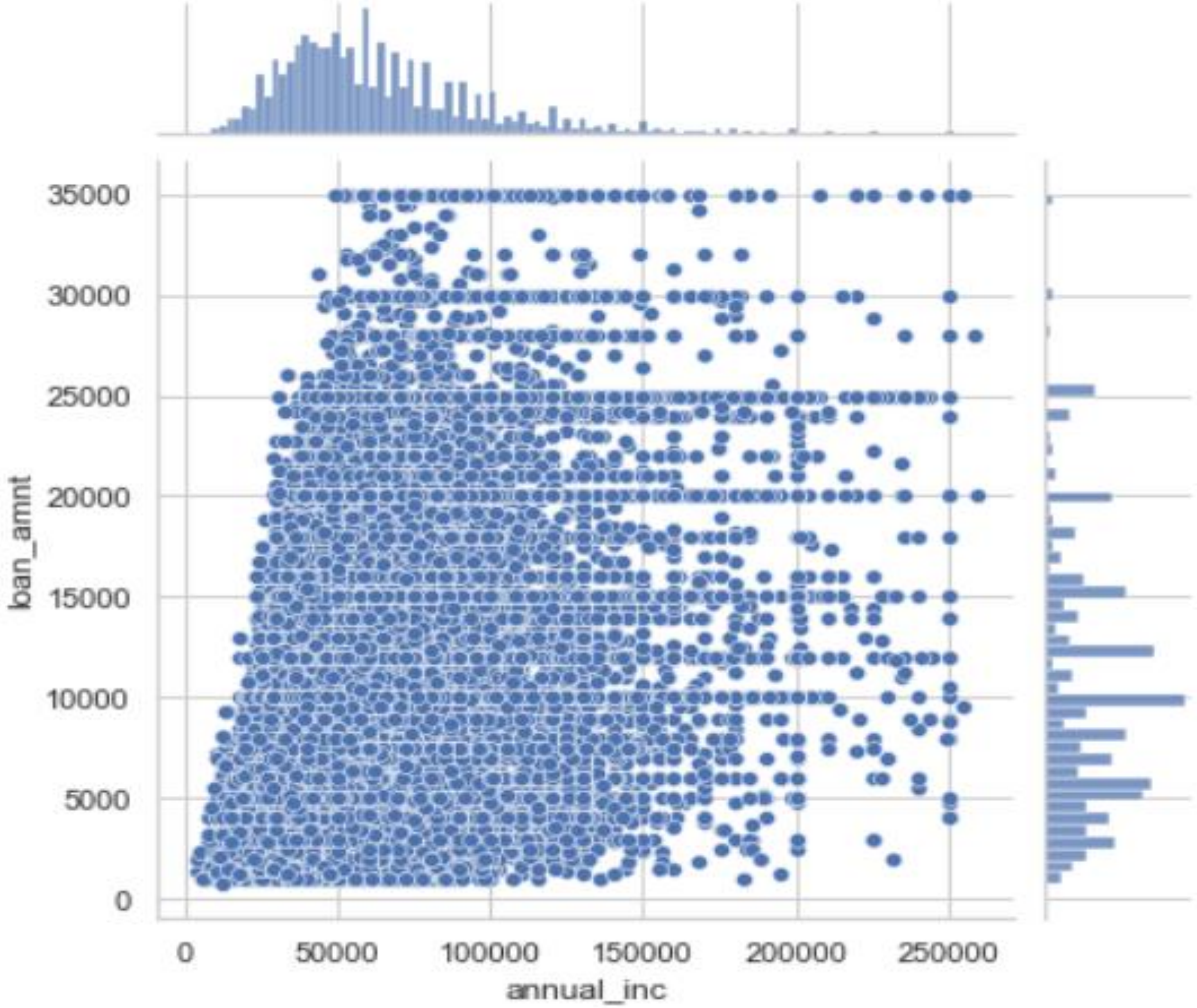


Analysis - Defaults by ratio of amount to income



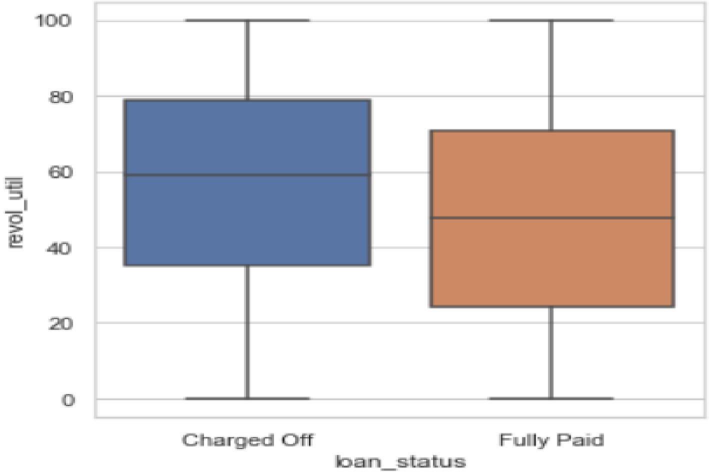
- As long as loan amount is less than 20% of annual income, defaults are low.
- Loan amounts of 30% of annual income or higher see a high rate of default.

Analysis - Defaults by ratio Continued



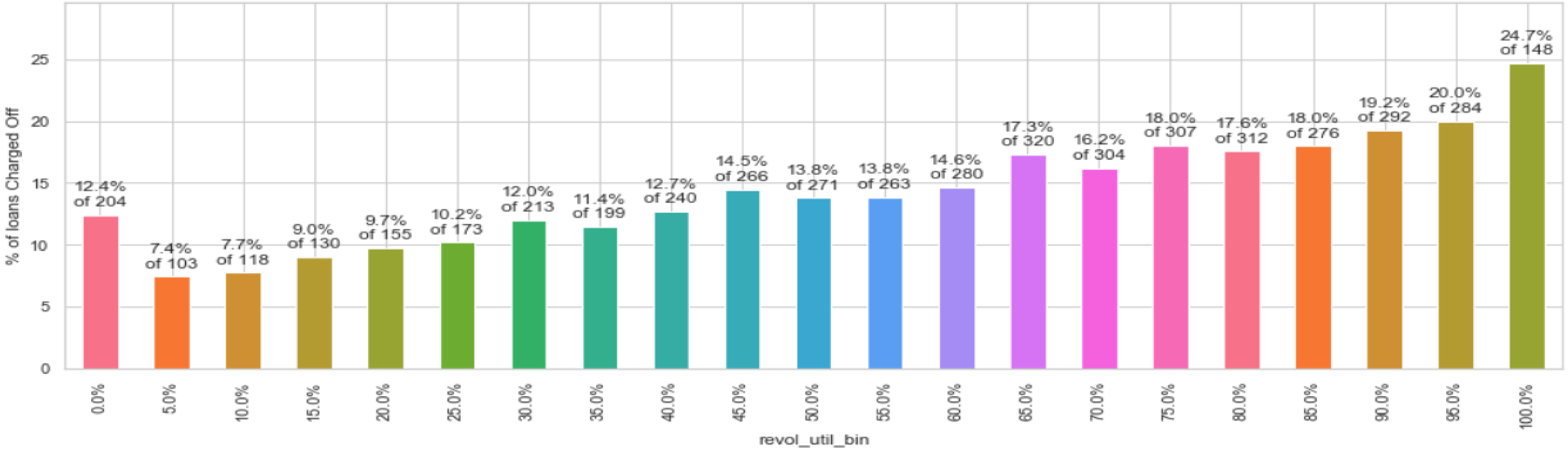
- We see here that Lending Club has extended high-value loans to people with low income.
- There are many cases of people with income 50000 or less getting loans of 25000 or more.

Analysis - Defaults by Revolving Line Util Rate

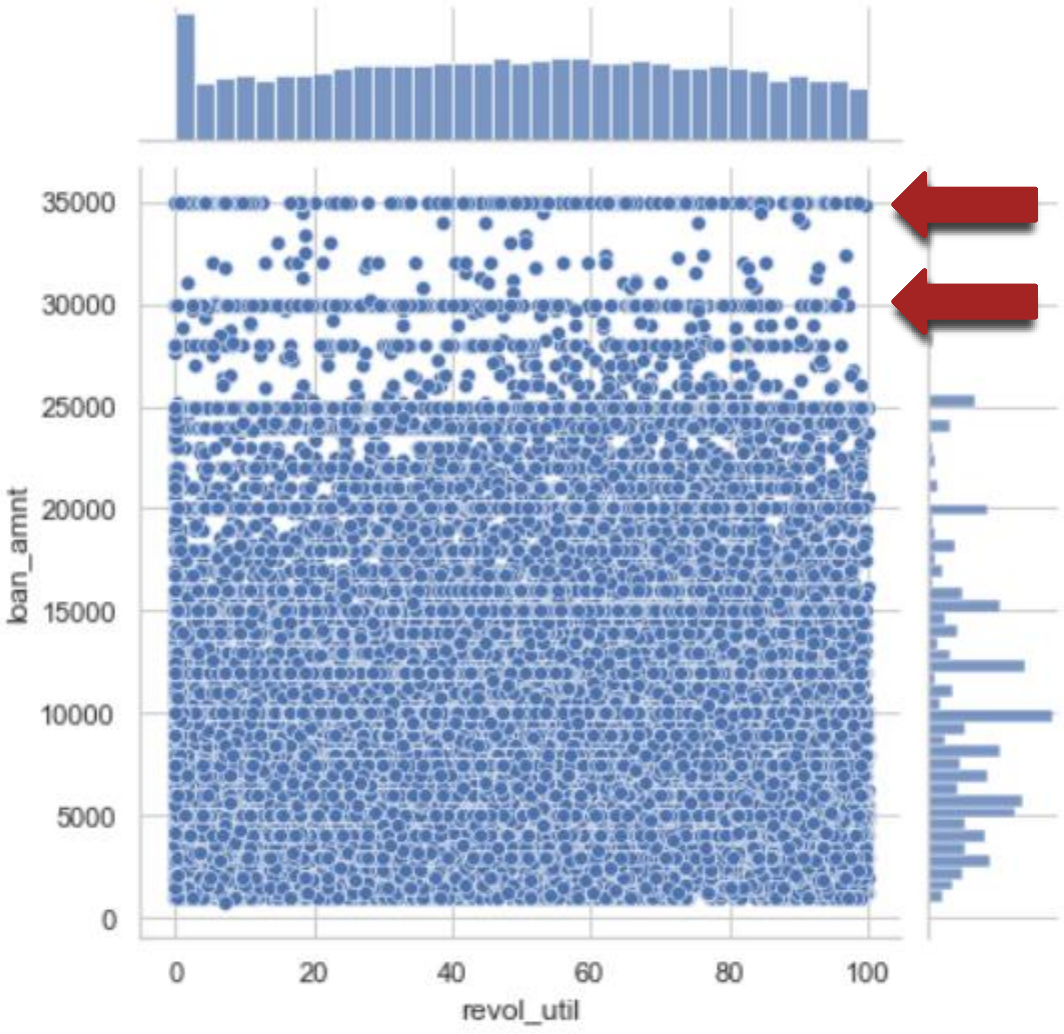


People with high utilization of Revolving Line of Credit at the time of taking loan default more.

Loans with utilization > 75% are risky.



Analysis - Defaults by Revolving Line Util Rate Continued



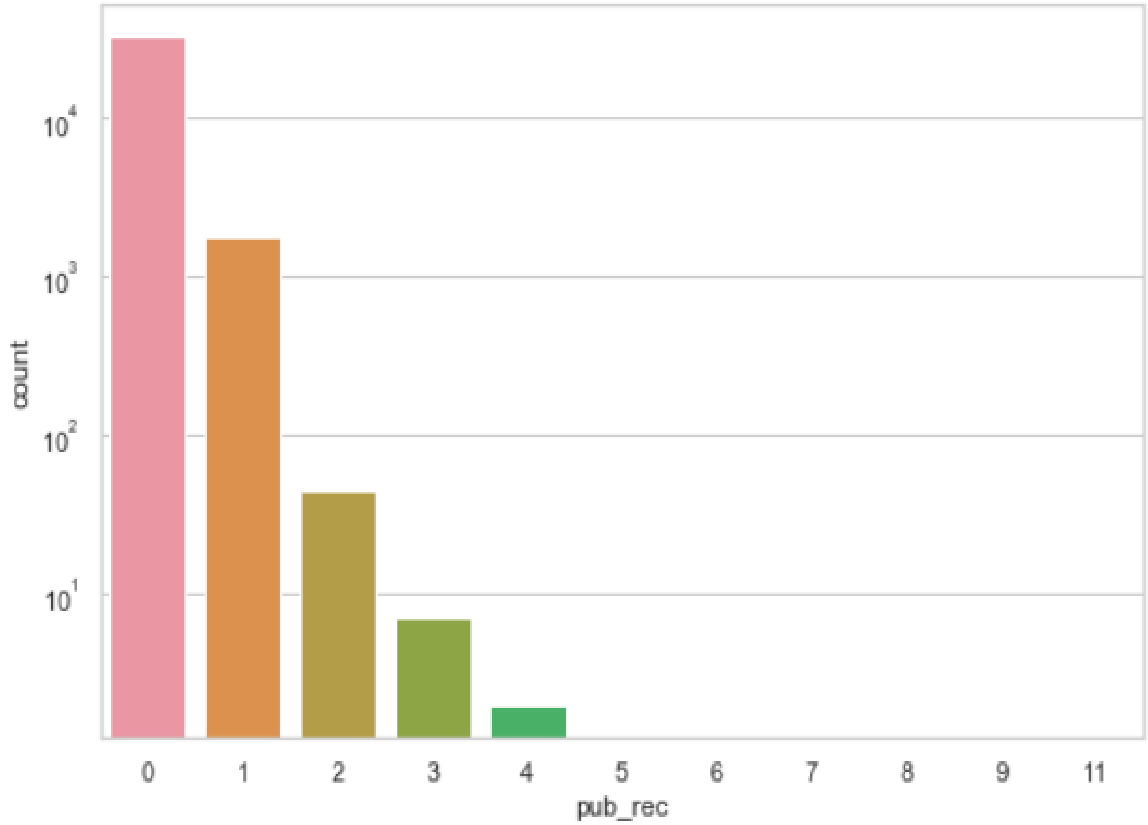
There have been some high value loans extended to borrowers with revolving line utilization rate of higher than 75%.

This practice should be stopped.

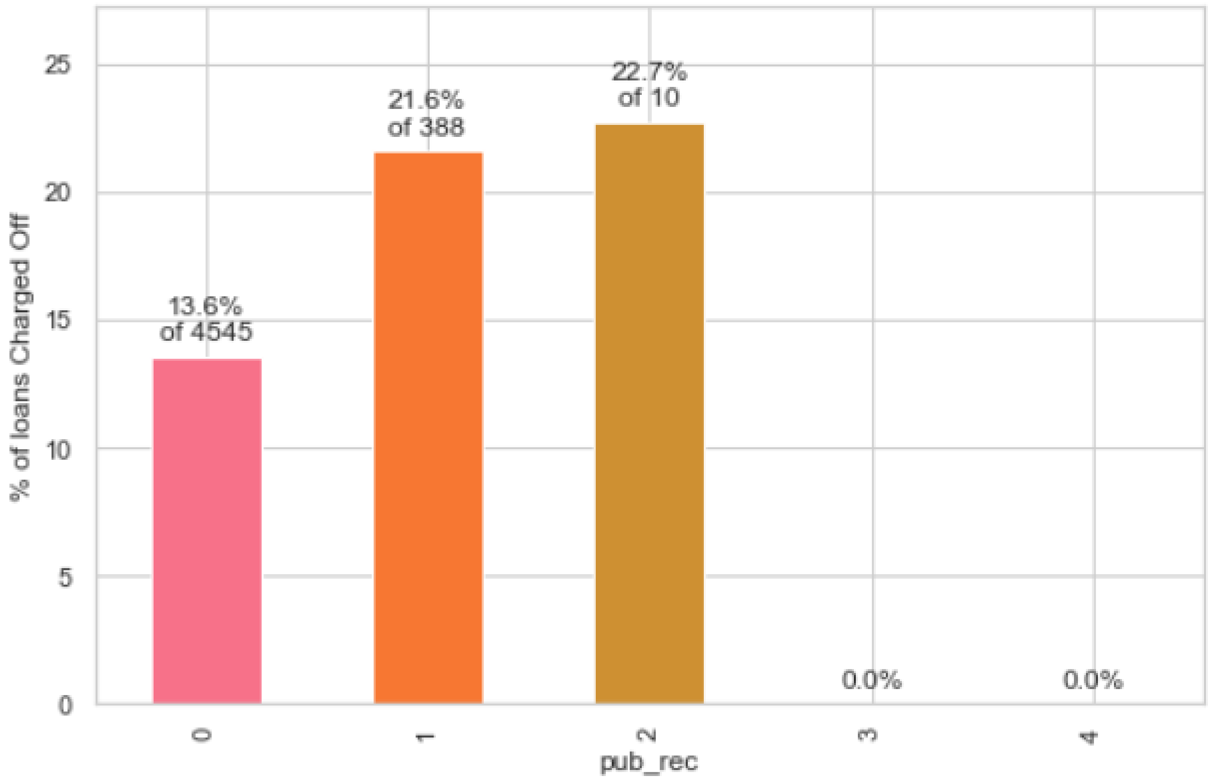
Density of low value loans is also high. They should be approved less often.

Analysis - Defaults by prior bad record

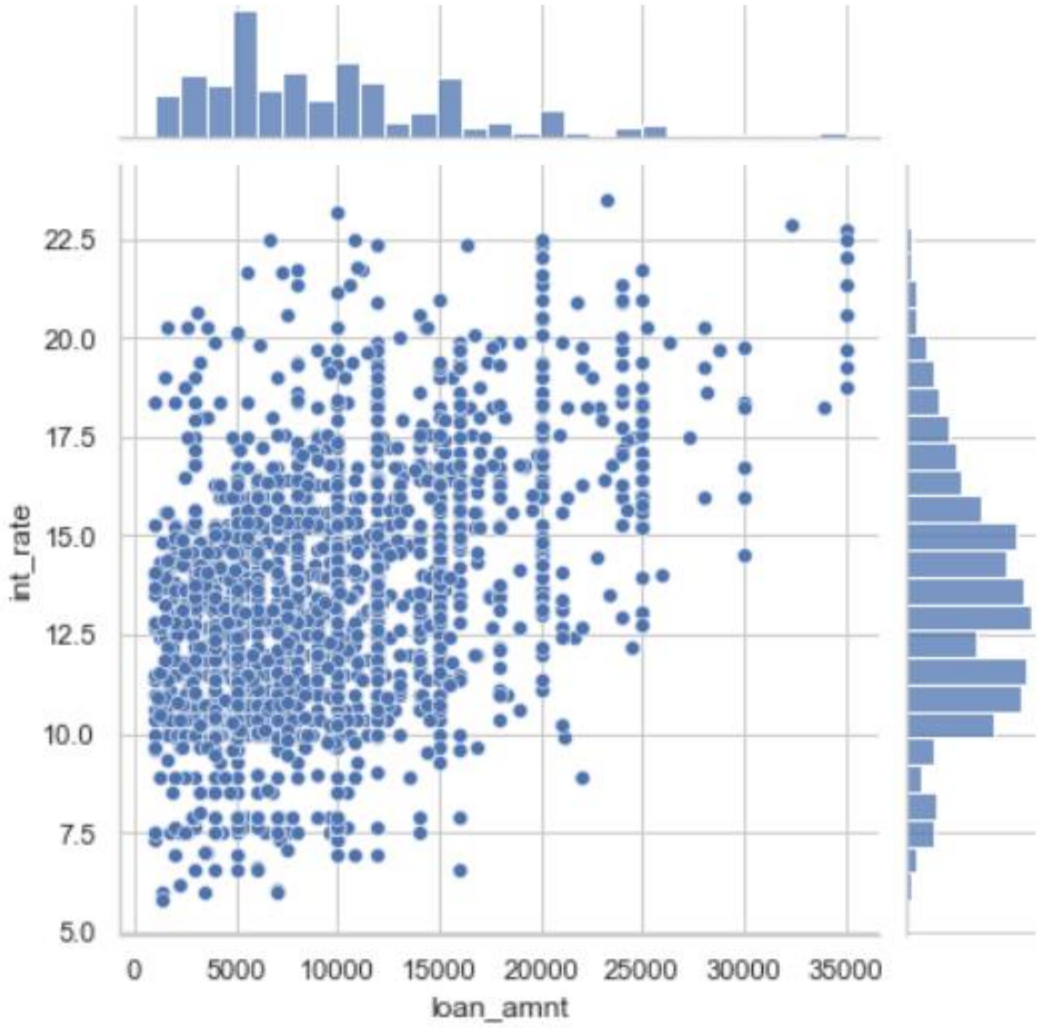
- 94% have no Public derogatory records.
- Having even 1 derogatory record increases the chances of Charge Off significantly.



- 96% have no bankruptcy record.
- Having even 1 bankruptcy record increases the chances of Charge Off significantly.
- Public Derogatory Record and Public Bankruptcy records have 83% correlation. We can use any one.



Analysis by prior bad record - Continued



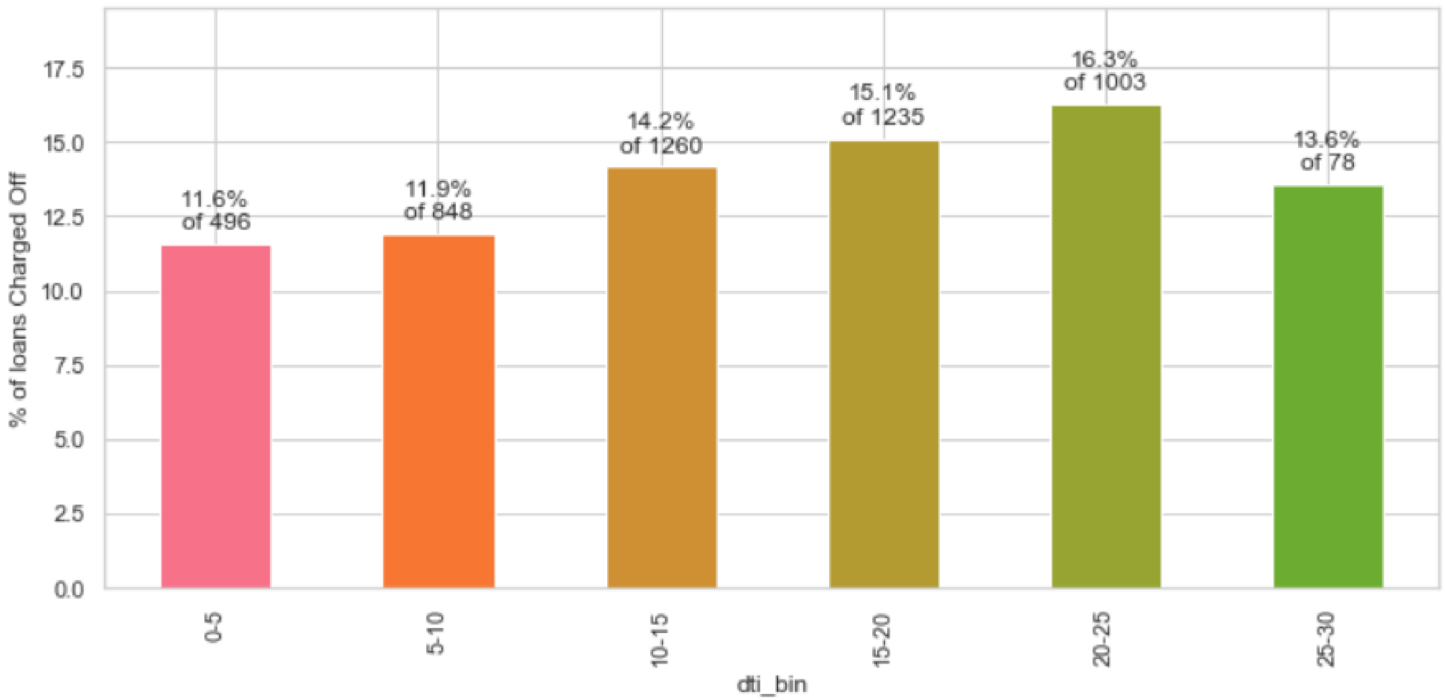
High value loans, as well as low interest loans have been extended to those with prior public derogatory records.

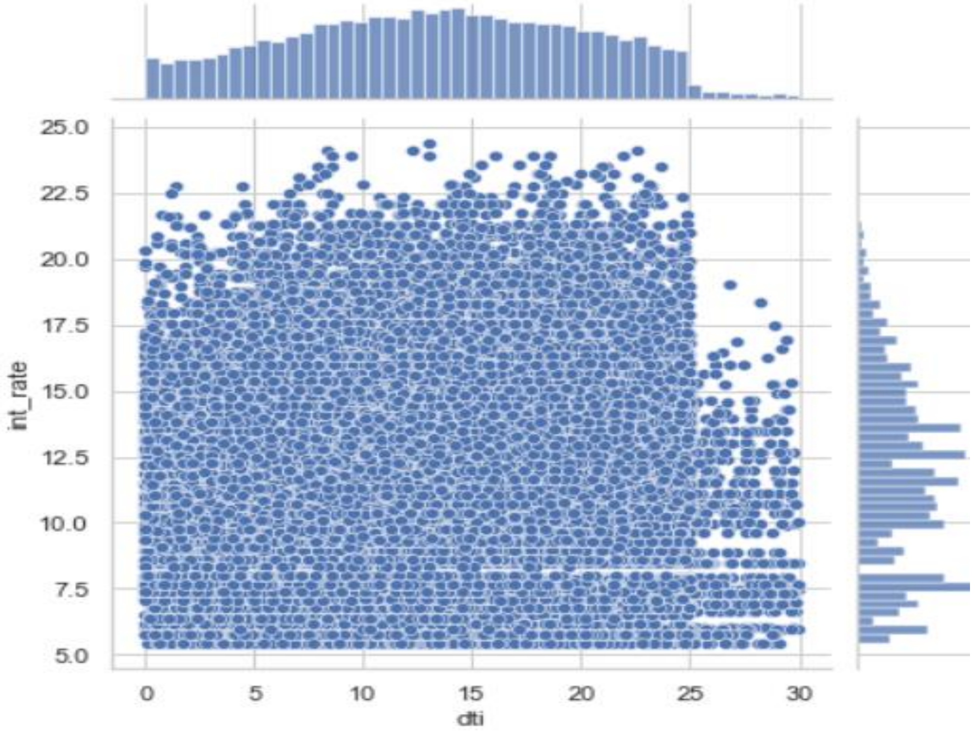
This practice can be stopped to improve business metrics.

Analysis – Defaults by Debt to Income Ratio

- Percentage of default rises with dti ratio.
- As the dti ratio rises above 20, the loans become risky.

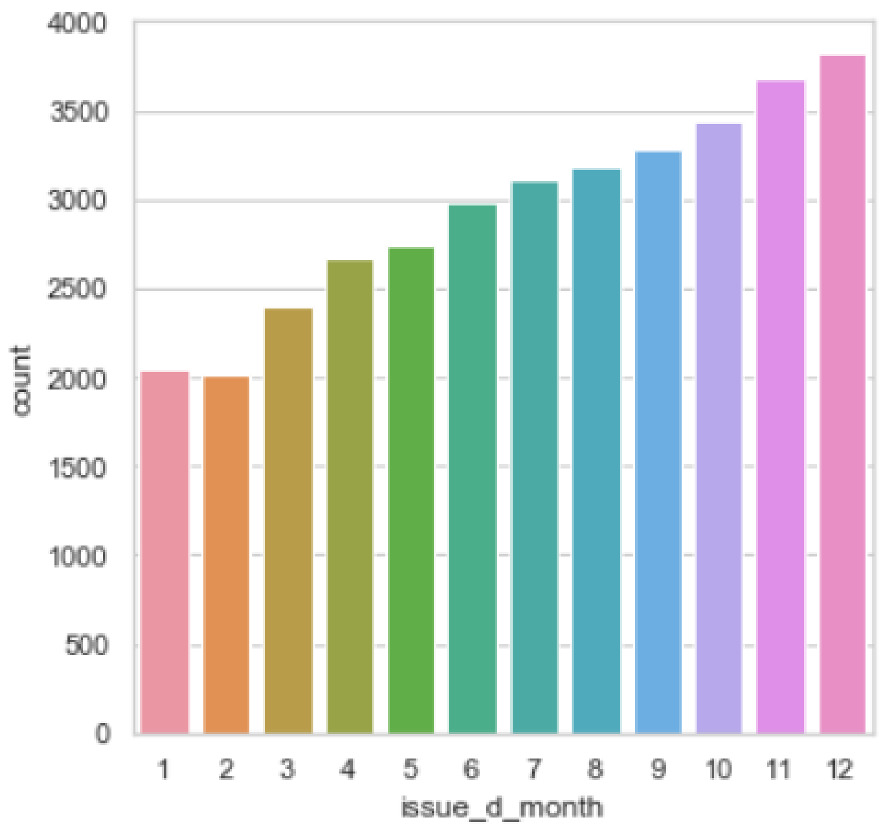
Higher interest rates should be charged for higher DTI, but we see spread across all values



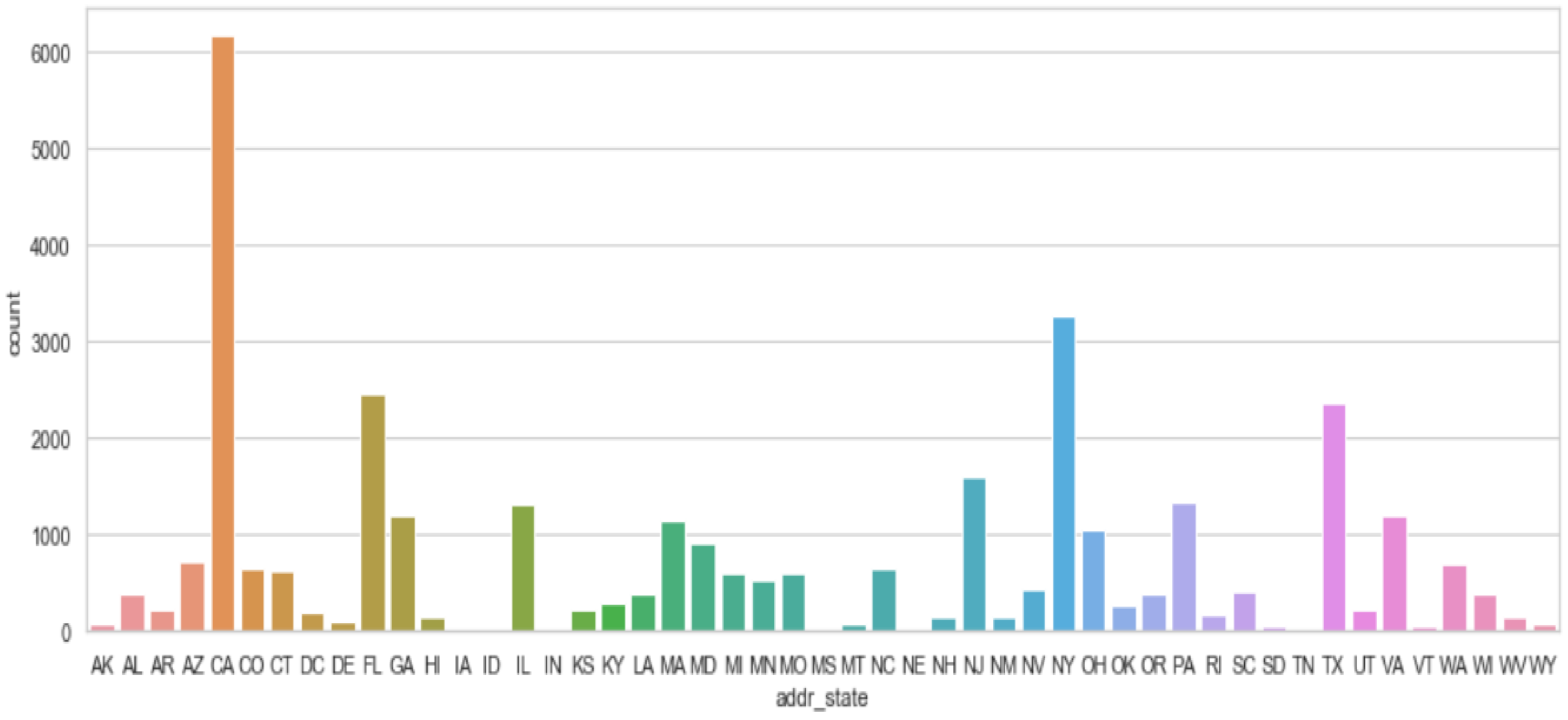


Some Interesting Observations

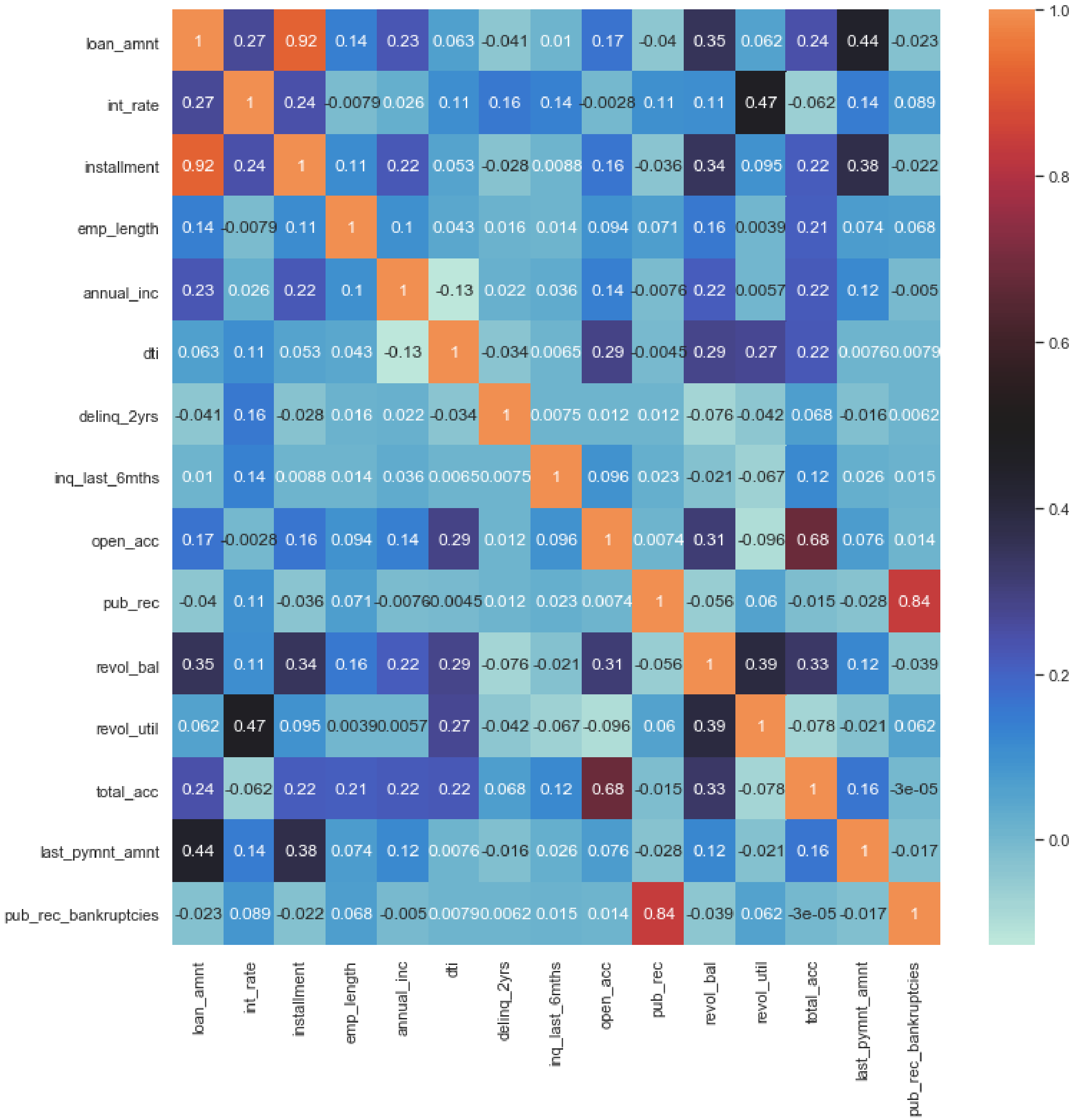
Loan numbers increase as months get closer to year-end.



Maximum loans are in populous states, California, New York, Florida and Texas



Correlation and Observation



Observation:

- Loan_amount is correlated to last_payment_amount with r factor.44, as expected
- int_rate is correlated to revol_util with r factor of .47 - This is good, as company is charging higher interest from riskier loan.
- loan_amnt revol_bal are correlated with r factor .35 - This is not good as it suggests that higher loan amount is being approved to riskier borrowers.</mark>
- delinq_2yrs is totally un-correlated with public record of bankruptcy. Therefore they represent distinct features with individual predictive value.

Conclusion:

Below are the top variables that are impacting the defaulters:

- **Grade, SubGrade** -- As grade changes there are defaulters changing, G being the highest
- **Purpose of the loan** -- Small business has more defaulters
- **Interest rate** -- As interest rate increases defaulters are moe
- **Term** -- There are more defaulters with 60 months term than 36 months
- **Annual income** -- If the annual income is more then high possibility that he pays the loan
- **Debt to Income** -- Percentage of default rises with dti ratio. As the dti ratio rises above 20, the loans become risky

Recommendations:

1. **Stop** – approving loans where amount/income is higher than 30%.
2. **Reduce** – number of approvals where purpose is small business.
3. **Stop** – approving high-value loans when revolving line utilization rate greater than 75%.
4. **Stop** – approving loans to people with prior bad record. Or at least stop approving high-value loans.
5. **Start** – charging higher interest rates for loans with dti greater than 20.