

# LENDING CLUB CASE STUDY

Rahul Kumar Dogra  
Sameer Kulkarni



# PROBLEM



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



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



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

### Clean Data

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

### Univariate Analysis

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

### Segmented Univariate Analysis

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

### Bivariate Analysis

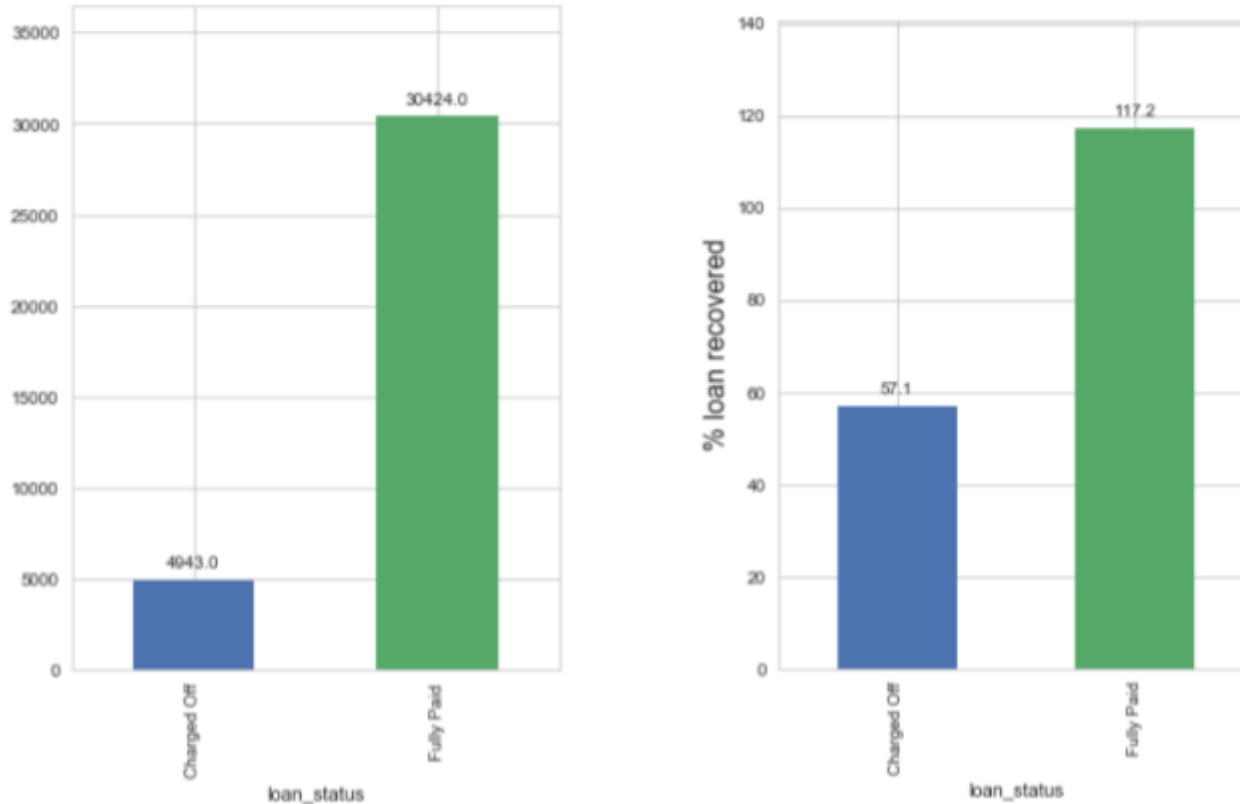
- Do correlation analysis
- Check how two variables affect each other or a third variable
- Analyze joint distributions

### Summarize Results

- Publish insights and observations

## ANALYSIS APPROACH

## Total Loans Vs Total Money Earned



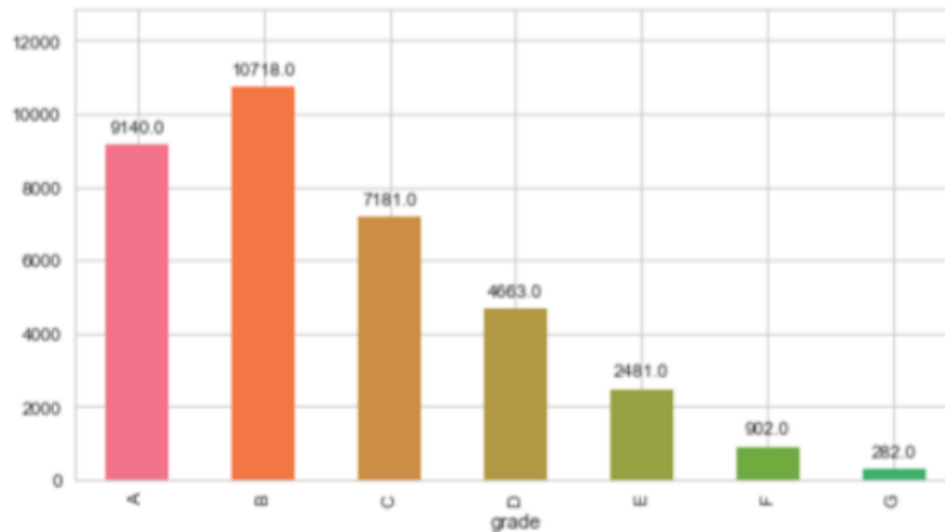
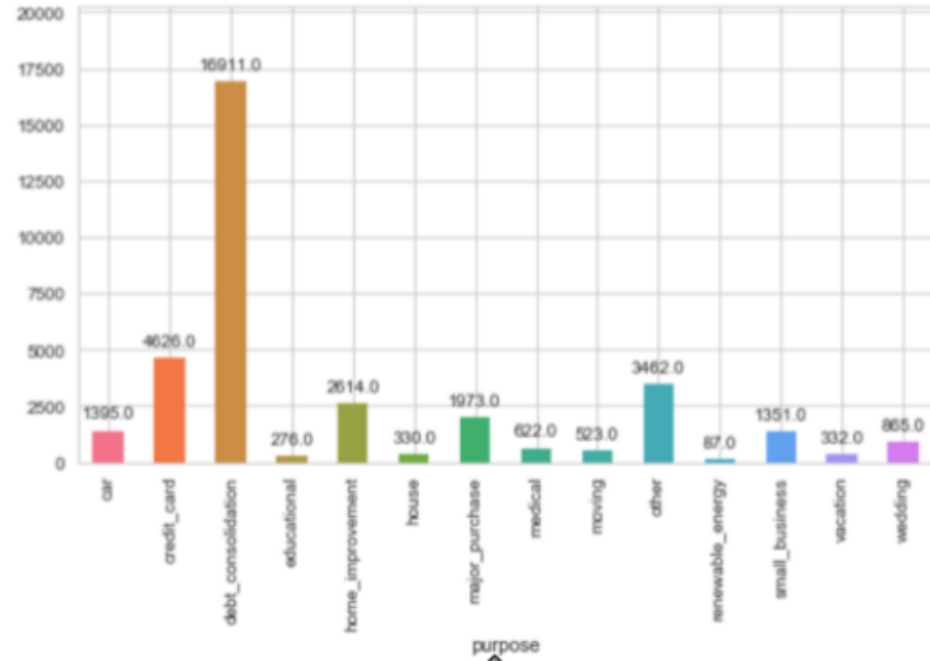
Approximately 14% of loans are defaulted Any variable that increases percentage of default to higher than 16.5% should be considered a business risk.

Lending Club only recovers 57% of the loan amount when loans are defaulted. On fully paid up loans, the company makes 17% profit.

## ANALYSIS - OVERALL LOAN STATUS

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

# ANALYSIS - UNDERSTANDING LOANS

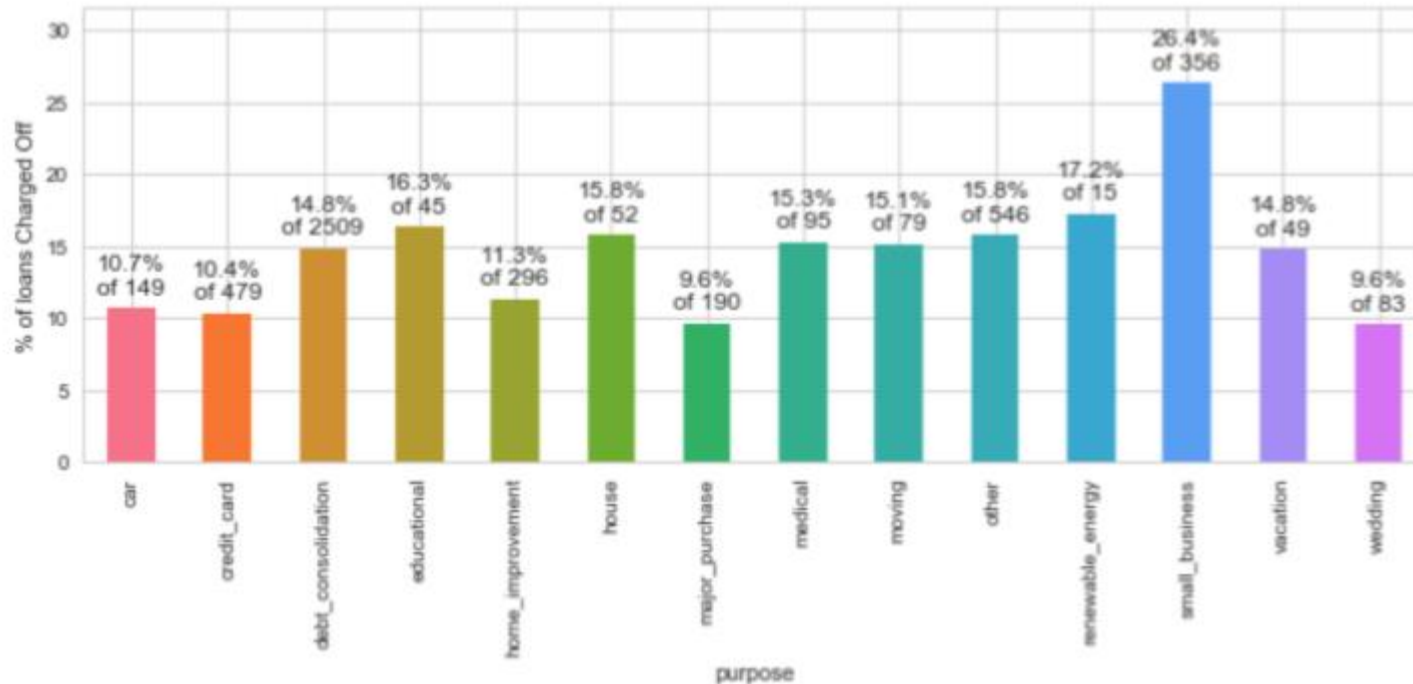
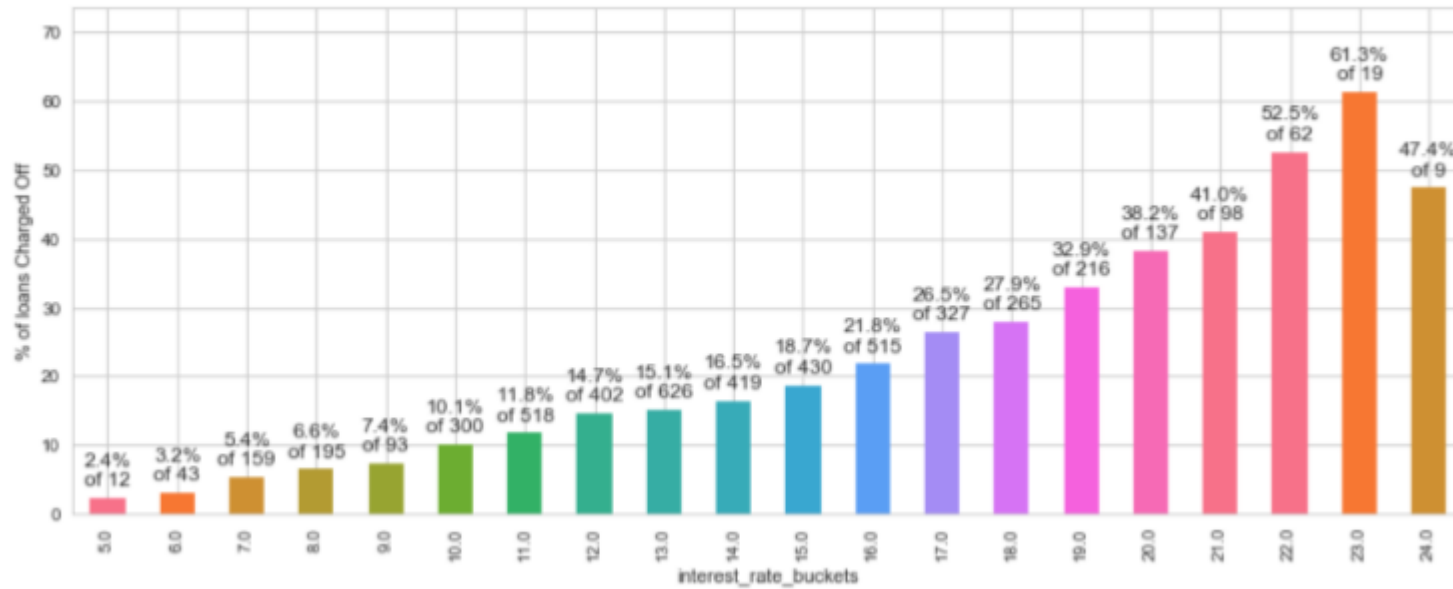


Maximum number of loans are for debt consolidation, followed by credit card.

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



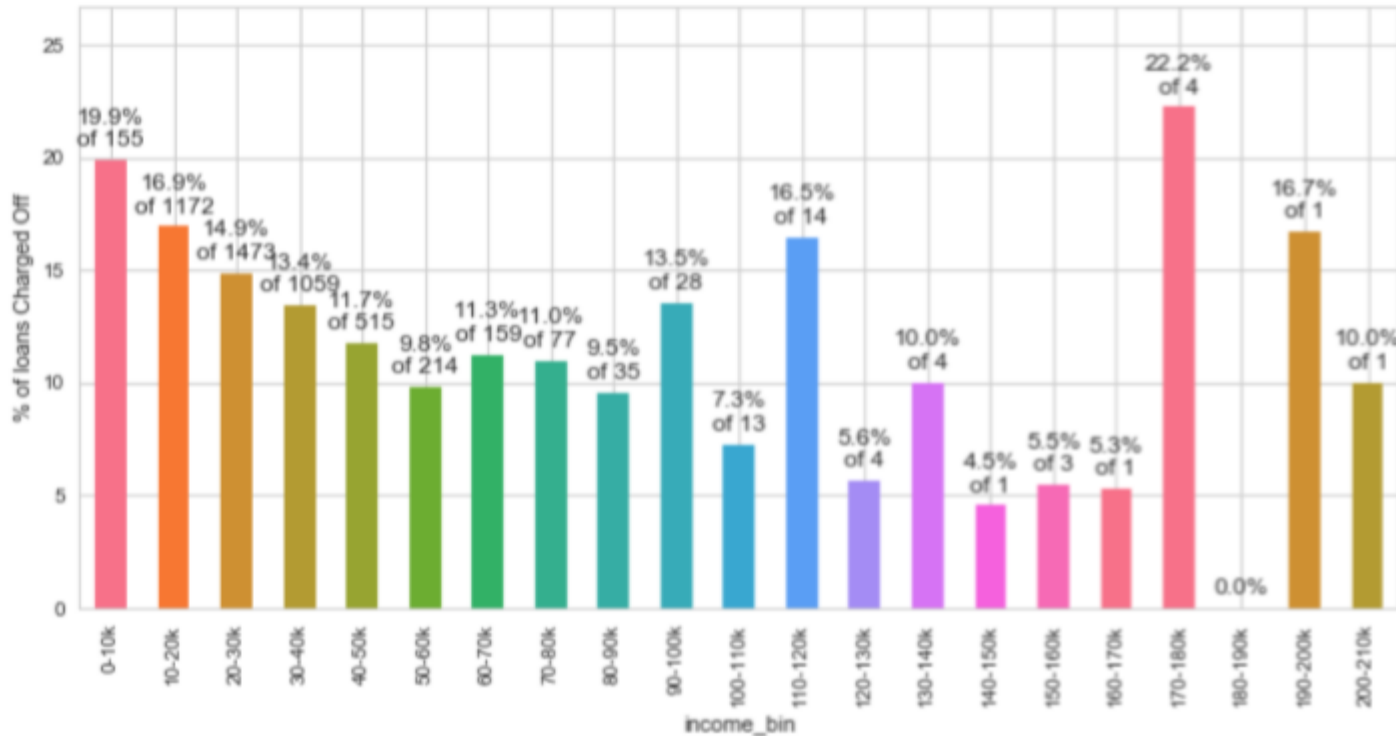
# ANALYSIS - DEFAULTS BY INTEREST RATE & LOAN PURPOSE



Percentage of Defaults increases monotonically with higher interest rates. At rates of 19% and above, more than 33% of loans are Charged Off.

More than a quarter of loans taken for the purpose of running a small business see defaults.

## 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.

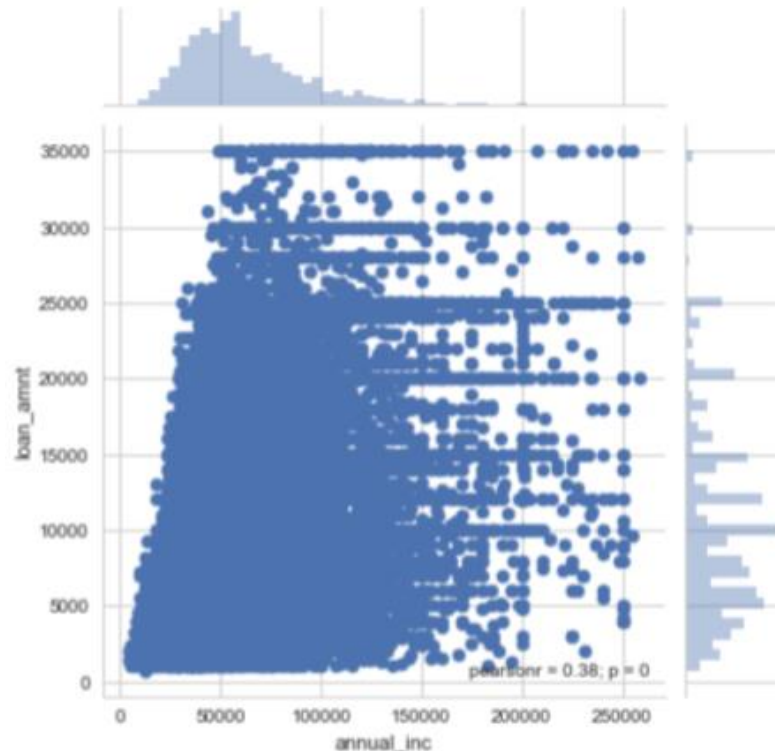
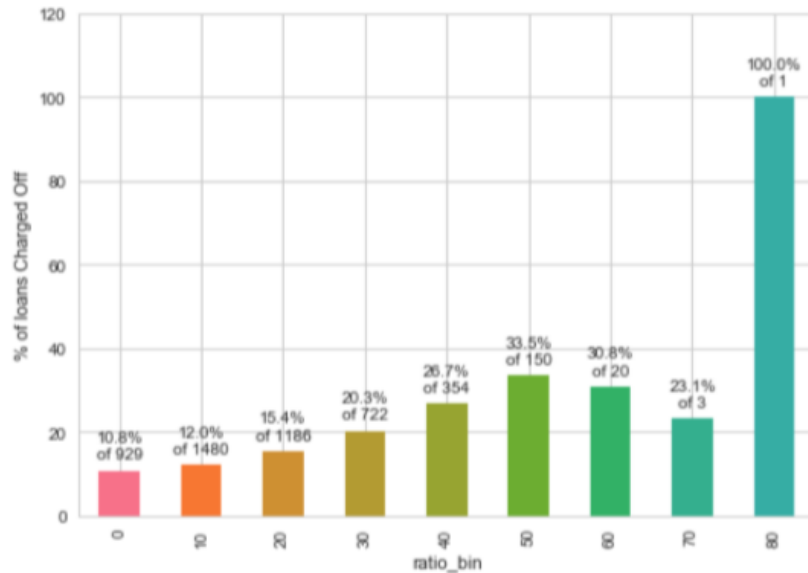
# ANALYSIS - DEFAULTS BY RATIO OF AMOUNT TO INCOME

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.

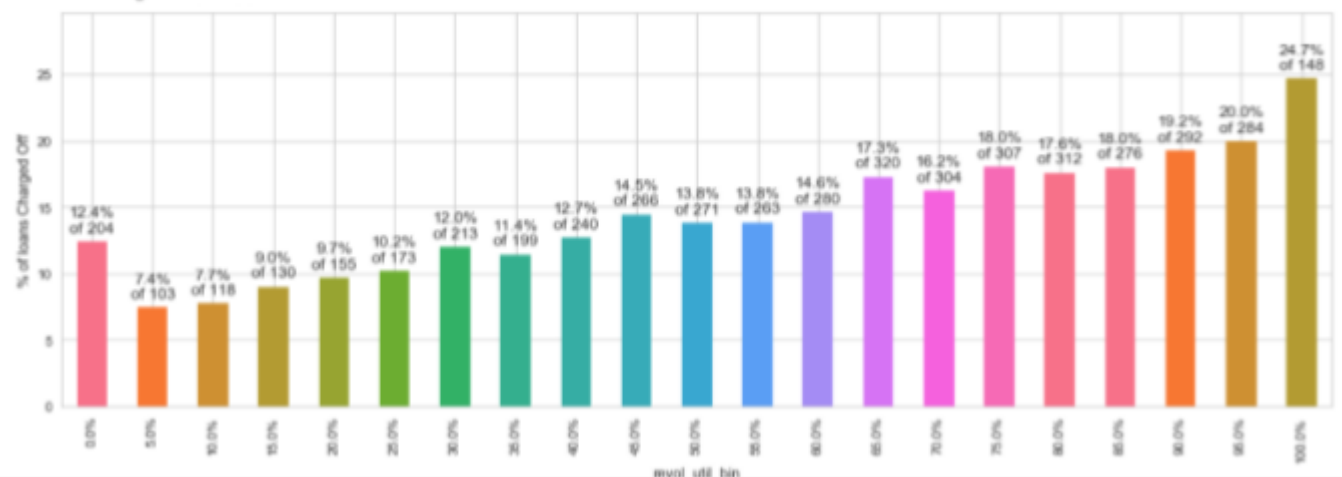
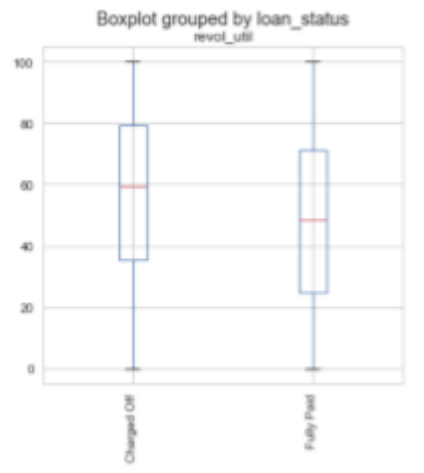
We see here that Lending Club has extended high-value loans to people with low income.

There are many cases of people with income 50000 or less getting loans of 25000 or more. This practice should be curtailed.



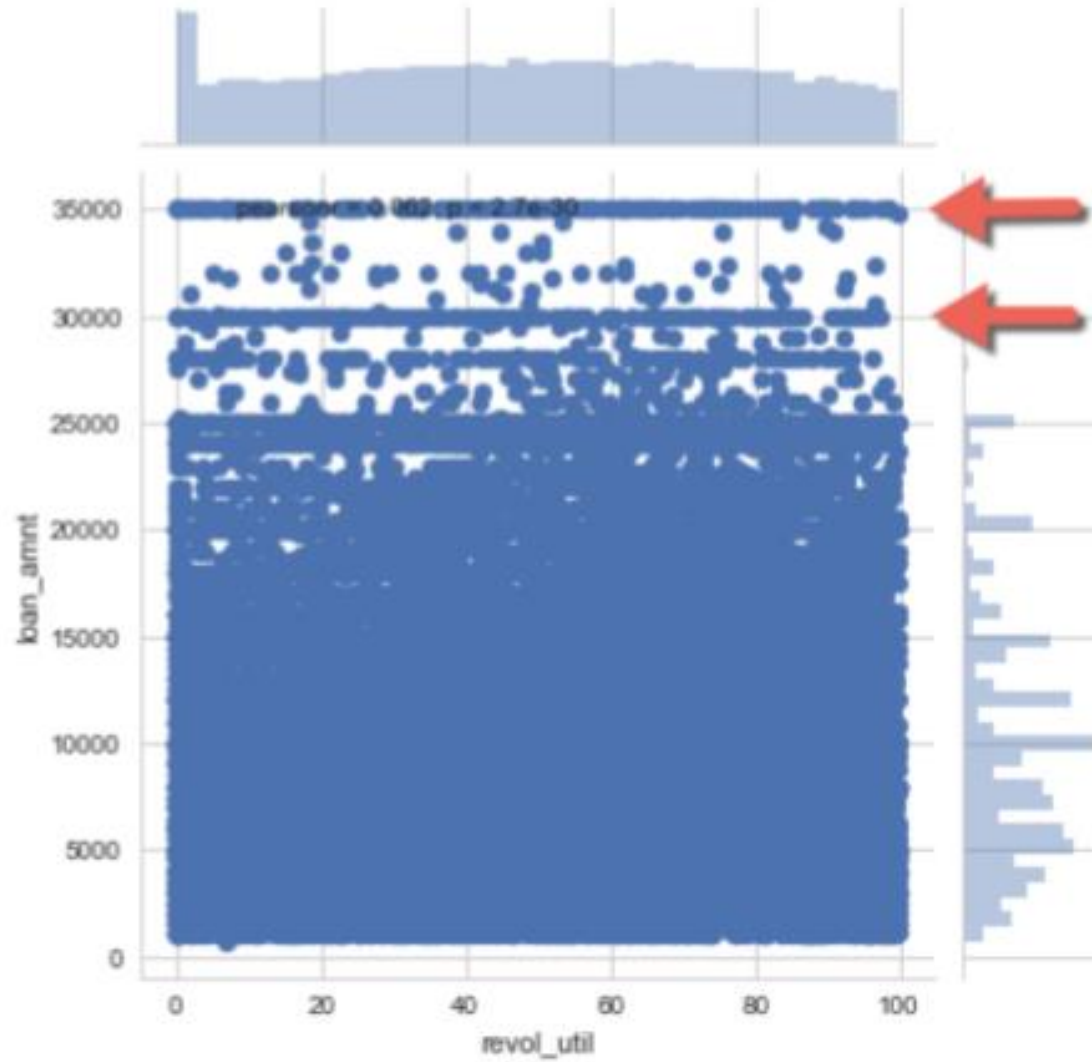


# ANALYSIS - DEFAULTS BY REVOLVING LINE UTIL RATE



People with high utilization of Revolving Line of Credit at the time of taking loan default more. Loans with utilization > 75% are risky.

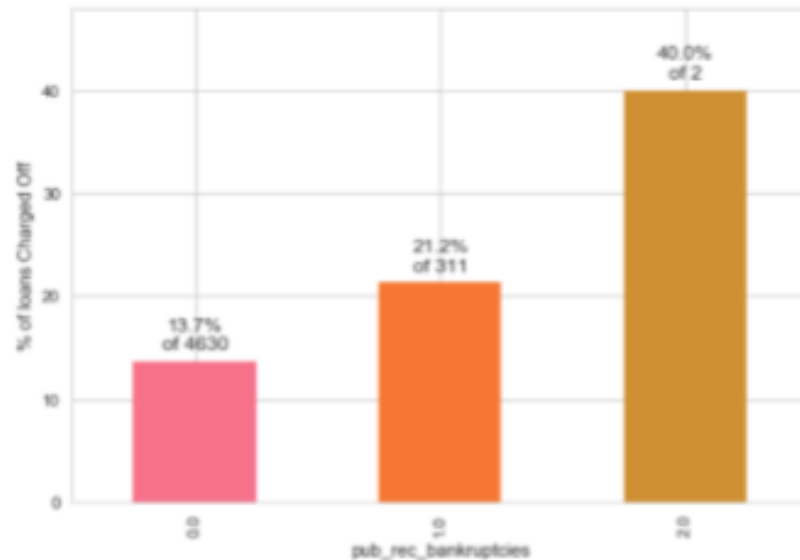
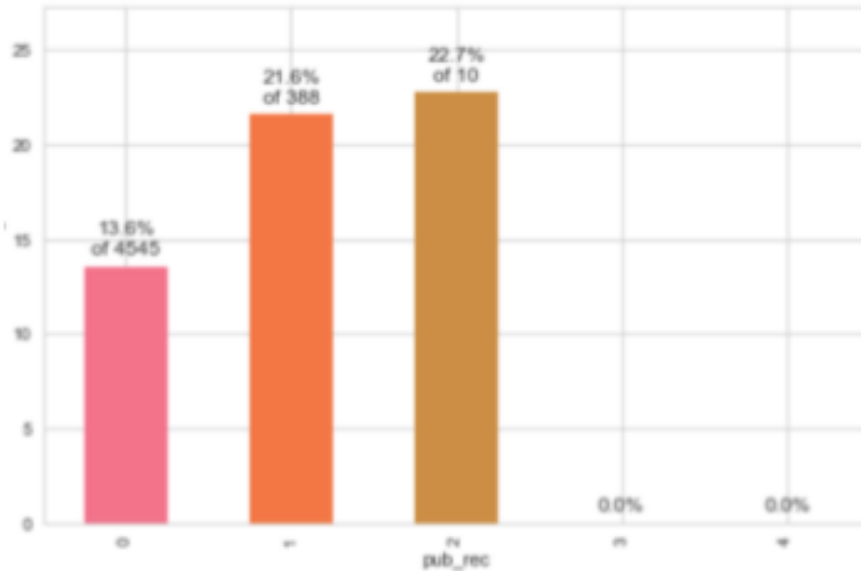
## ANALYSIS - DEFAULTS BY REV UTIL CONTINUED



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



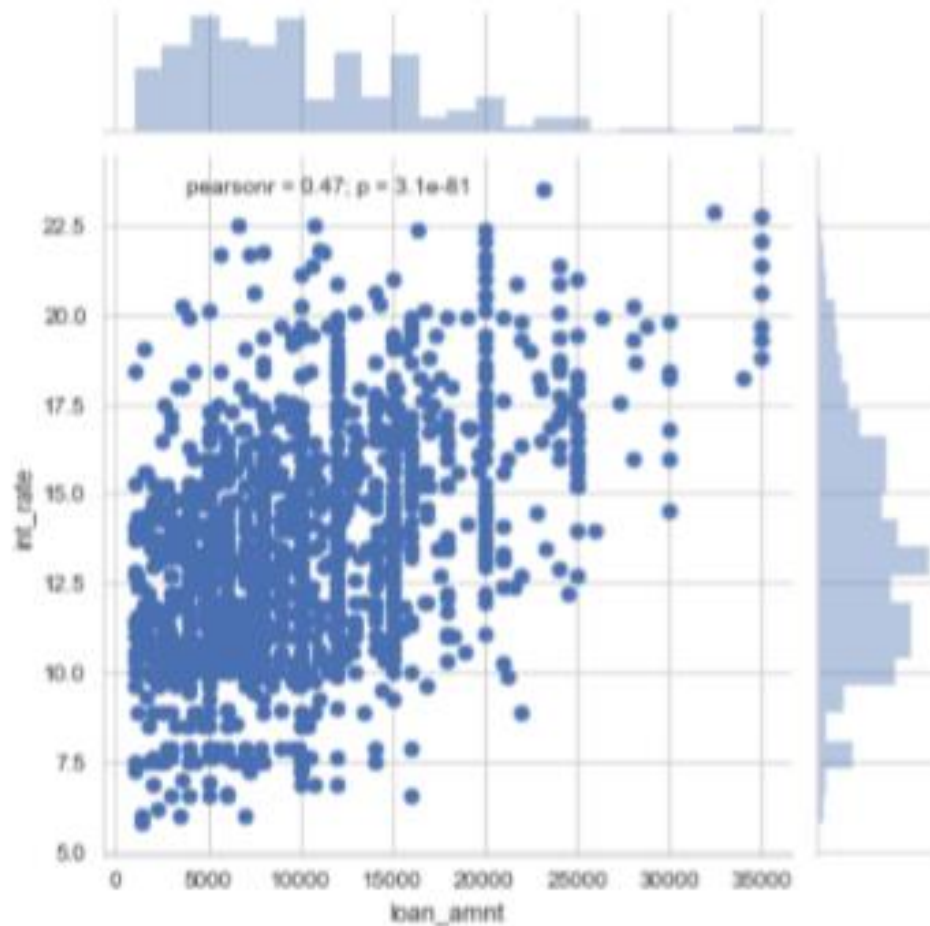
94% have no Public derogatory records.

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



Data of people with >0 bad records

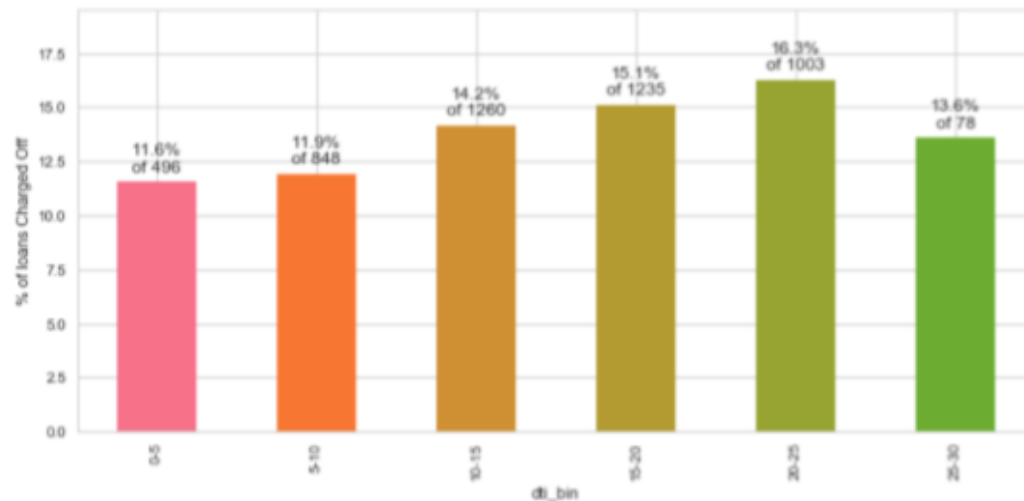
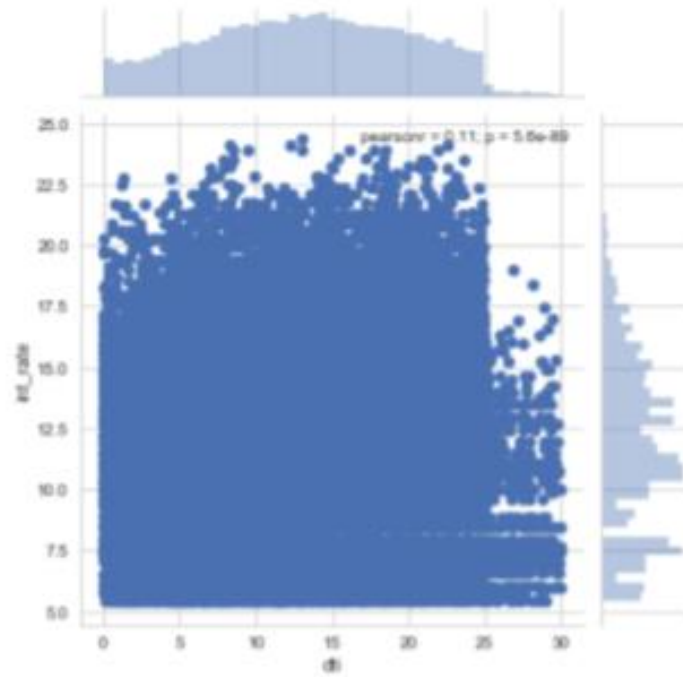
High value loans, as well as low interest loans have been extended to those with prior public derogatory records.

This practice can be stopped to improve business metrics.

# ANALYSIS – DEFAULTS BY DEBT TO INCOME RATIO

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

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

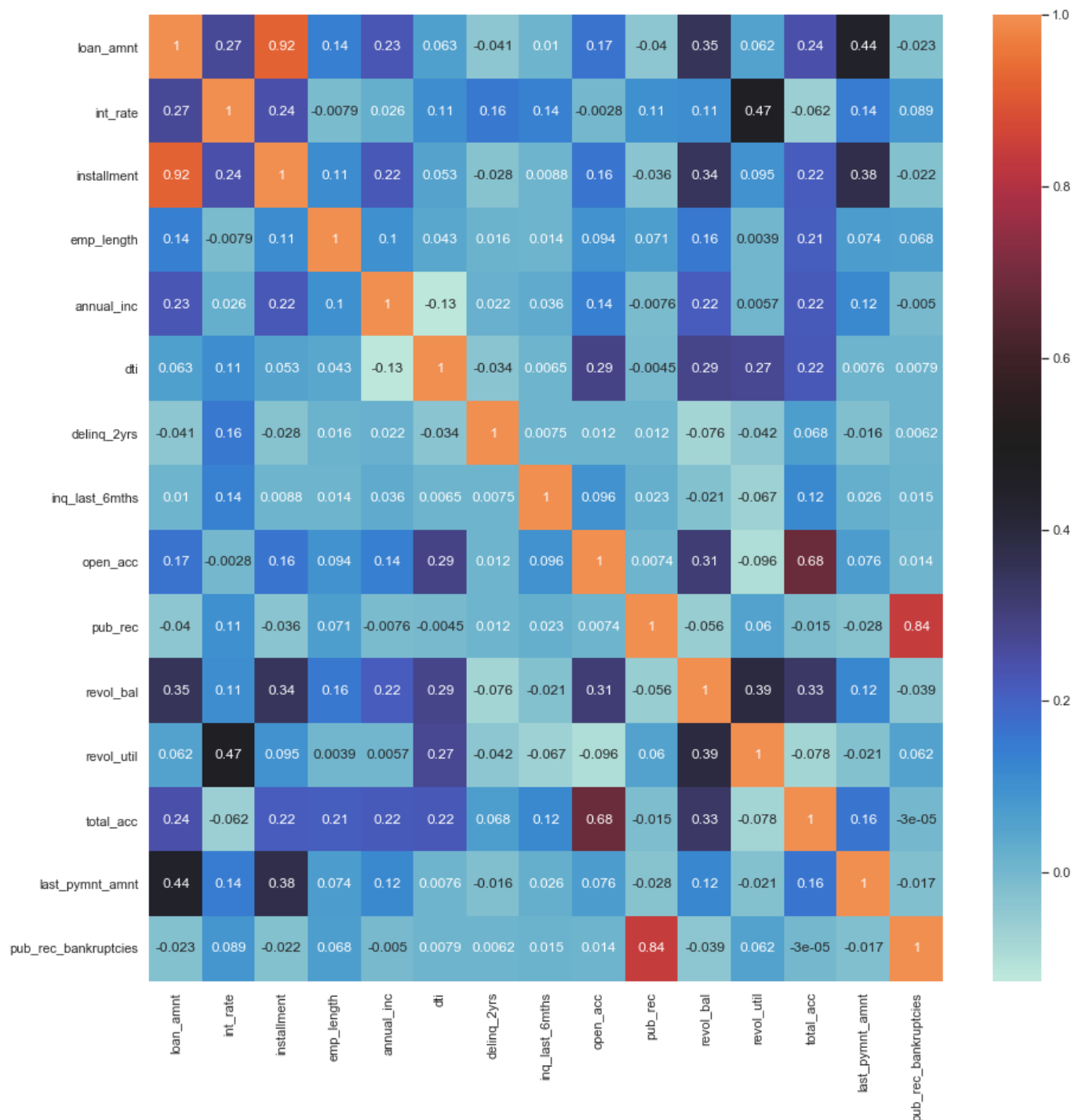


1. Stop –approving loans where amount/income is higher than 30%
2. Reduce –number of approvals where purpose is small business
3. Stop –approving high-value loans when revolving line utilization rate greater than 75%
4. Stop–approving loans to people with prior bad record. Or at least stop approving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

## RECOMMENDATIONS



# ADDITIONAL SLIDE (CORRELATIONS)



loan\_amnt 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.

delinq\_2yrs is totally un-correlated with public record of bankruptcy. Therefore they represent distinct features with individual predictive value.



THANK YOU

Rahul Kumar Dogra

Sameer Kulkarni