

Lending Club Case Study

Prabhat
Kumar
Prabhakar

Shravan
Kumar Jha



Problem Statement



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.



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



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

Approach



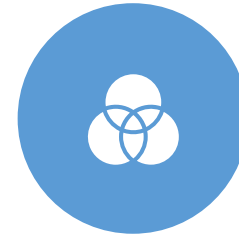
CLEAN DATA : DROP COLUMNS WITH NULL VALUES, DROP ROWS FOR VERY LESS MISSING VALUE COLUMNS. CONVERT FEATURES DATATYPES IN TO PROPER INT, FLOAT, DATE REPRESENTATIONS



UNIVARIATE ANALYSIS: CHECK DISTRIBUTIONS AND FREQUENCIES OF VARIOUS NUMERICAL AND CATEGORICAL VARIABLES



SEGMENTED UNIVARIATE ANALYSIS :ANALYZE VARIABLES AGAINST SEGMENTS OF OTHER VARIABLES



BI-VARIATE ANALYSIS: DO CORRELATION ANALYSIS. CHECK HOW TWO VARIABLES AFFECT EACH OTHER OR A THIRD VARIABLE



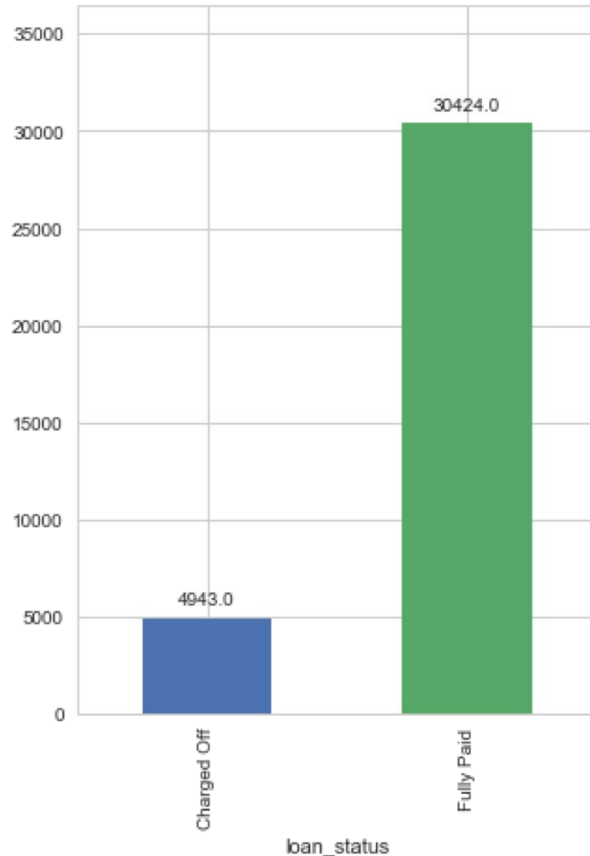
SUMMARIZE RESULTS: PUBLISH INSIGHTS AND OBSERVATIONS

Analysis - Overall Loan Status

Total Loans

Approximately **14%** of loans are defaulted

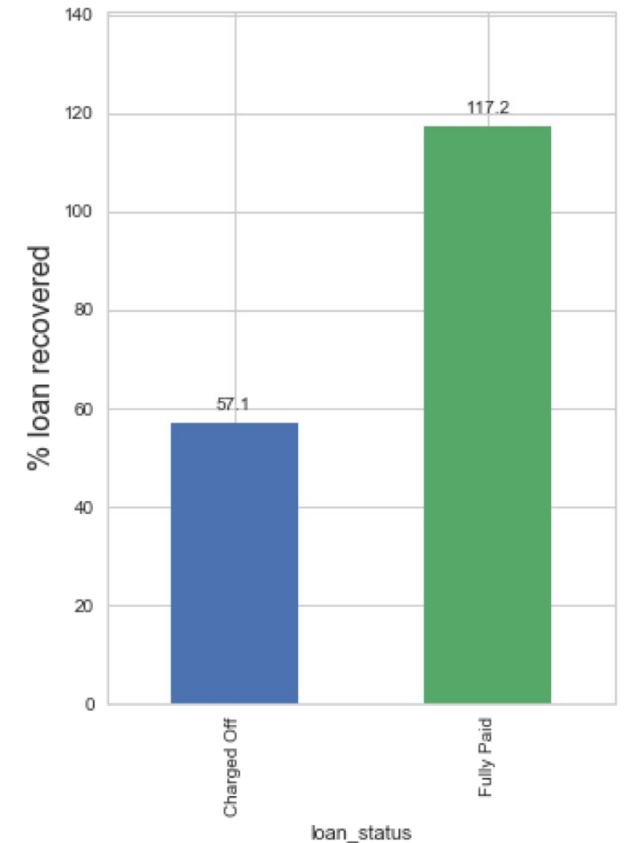
Any variable that increases percentage of default to higher than 16.5% should be considered a business risk.



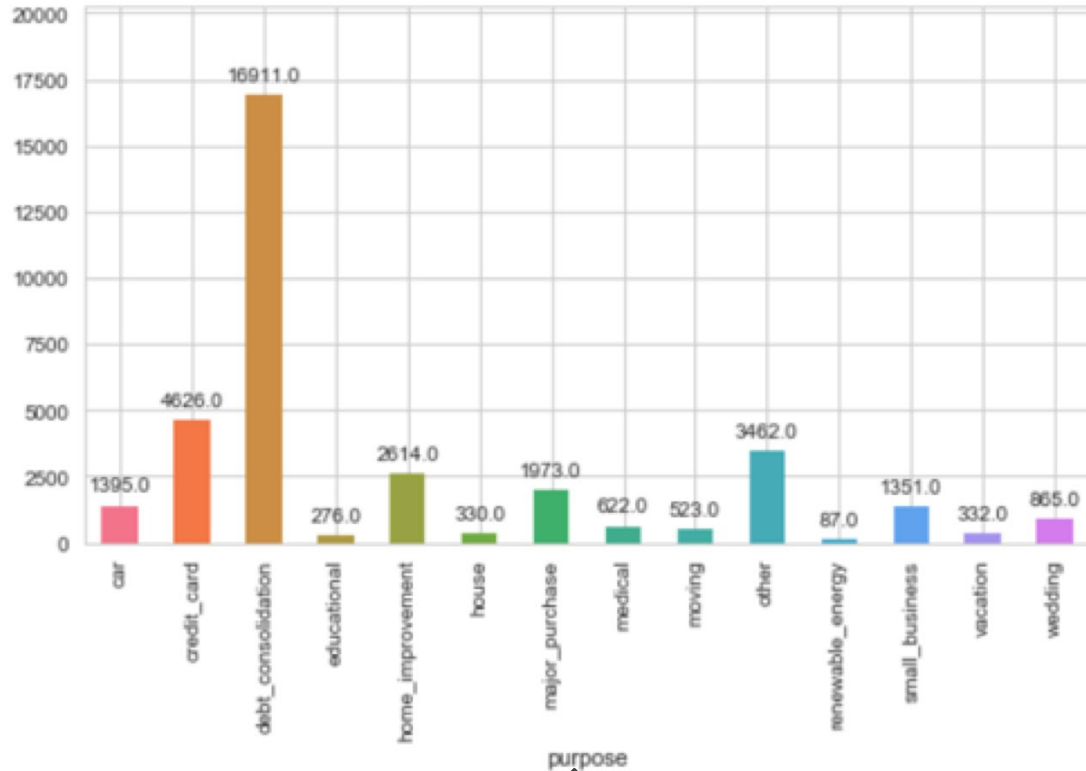
Total Money Earned

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

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

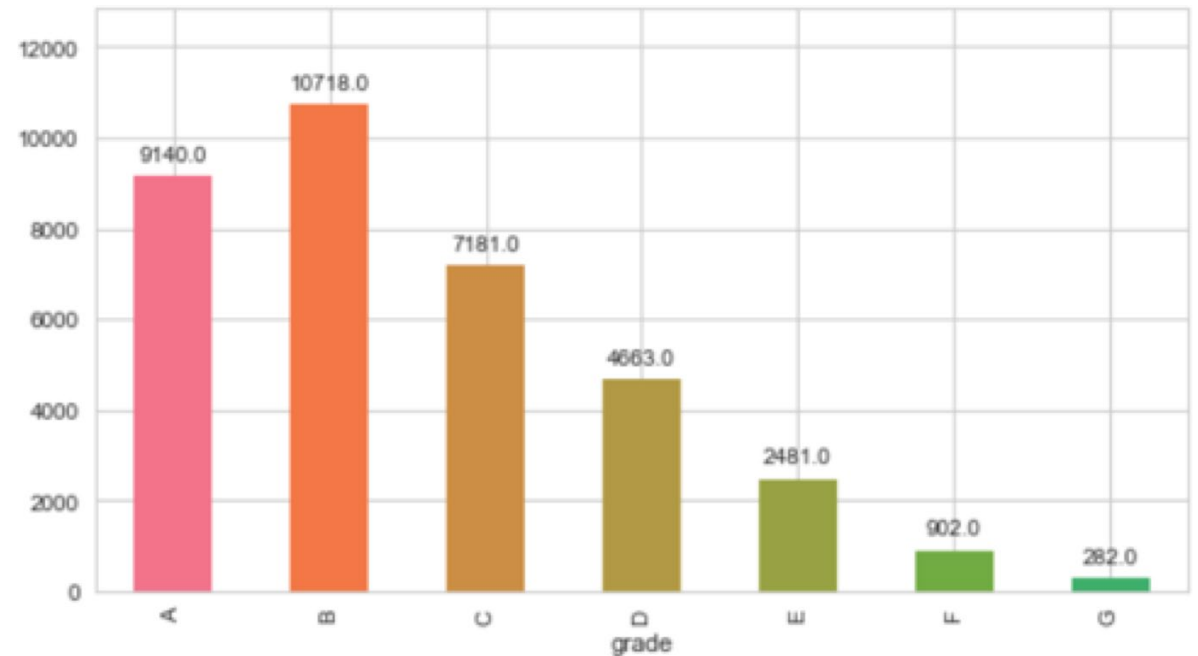


Analysis - Understanding



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

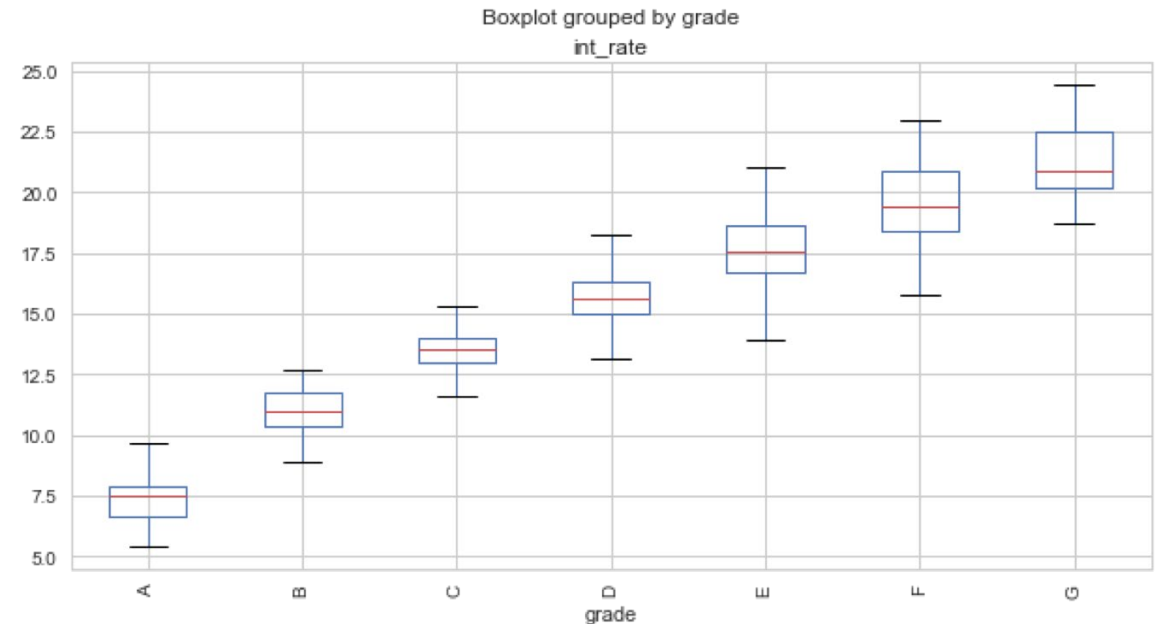
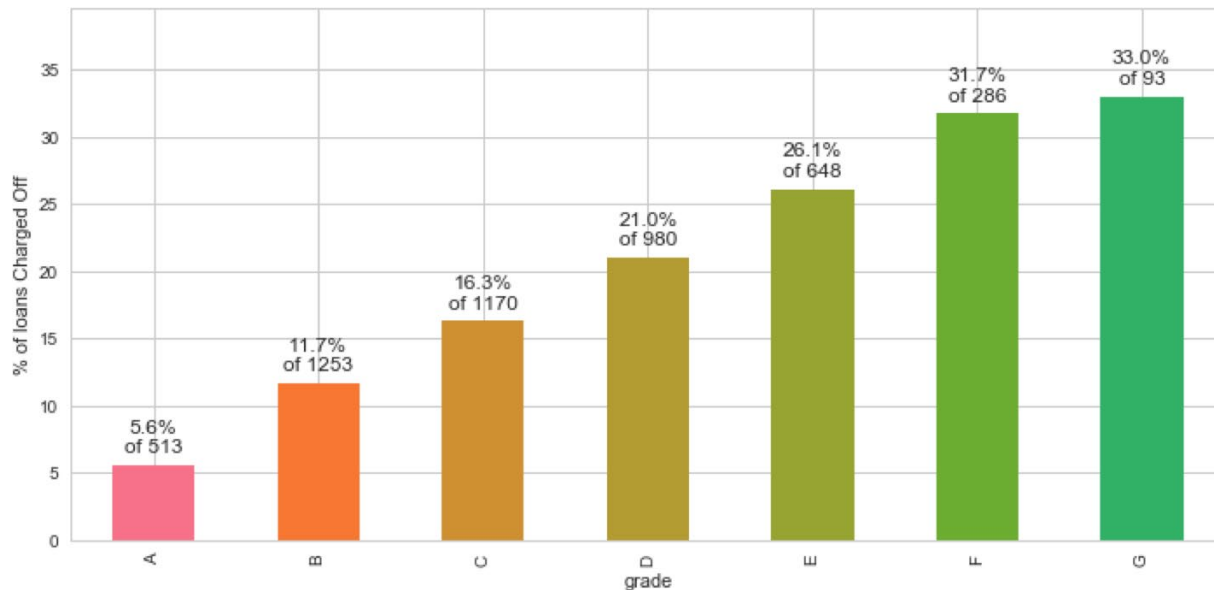
Most loans are high quality, with a grade of A or B



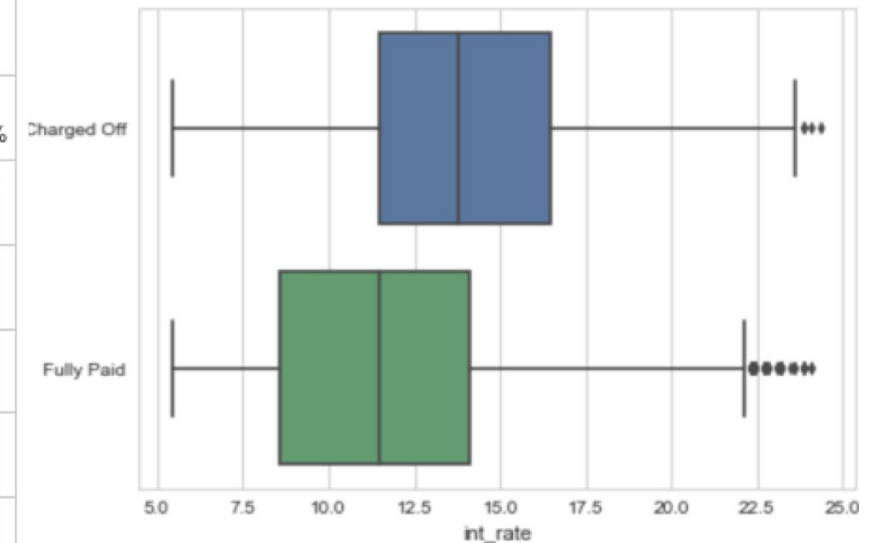
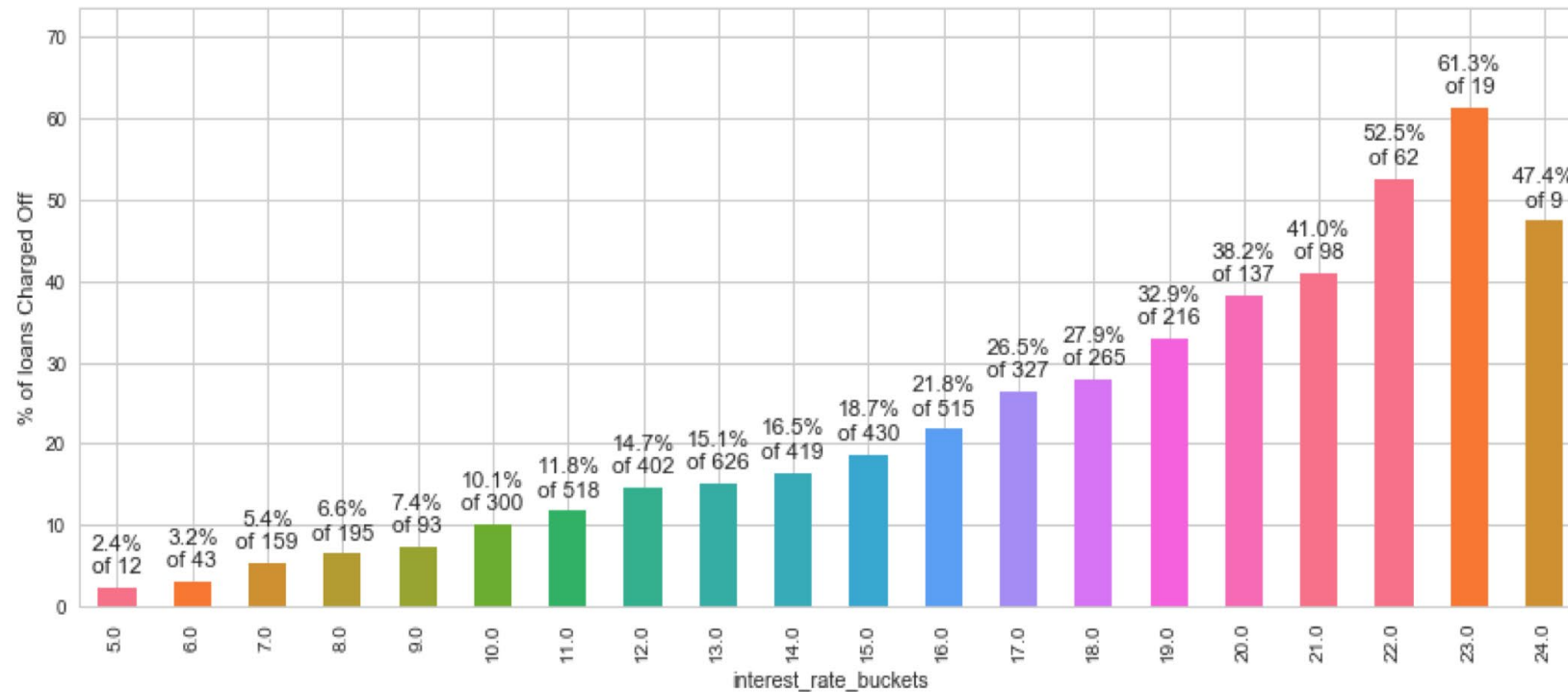
Analysis - Understanding Loans Continued

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.

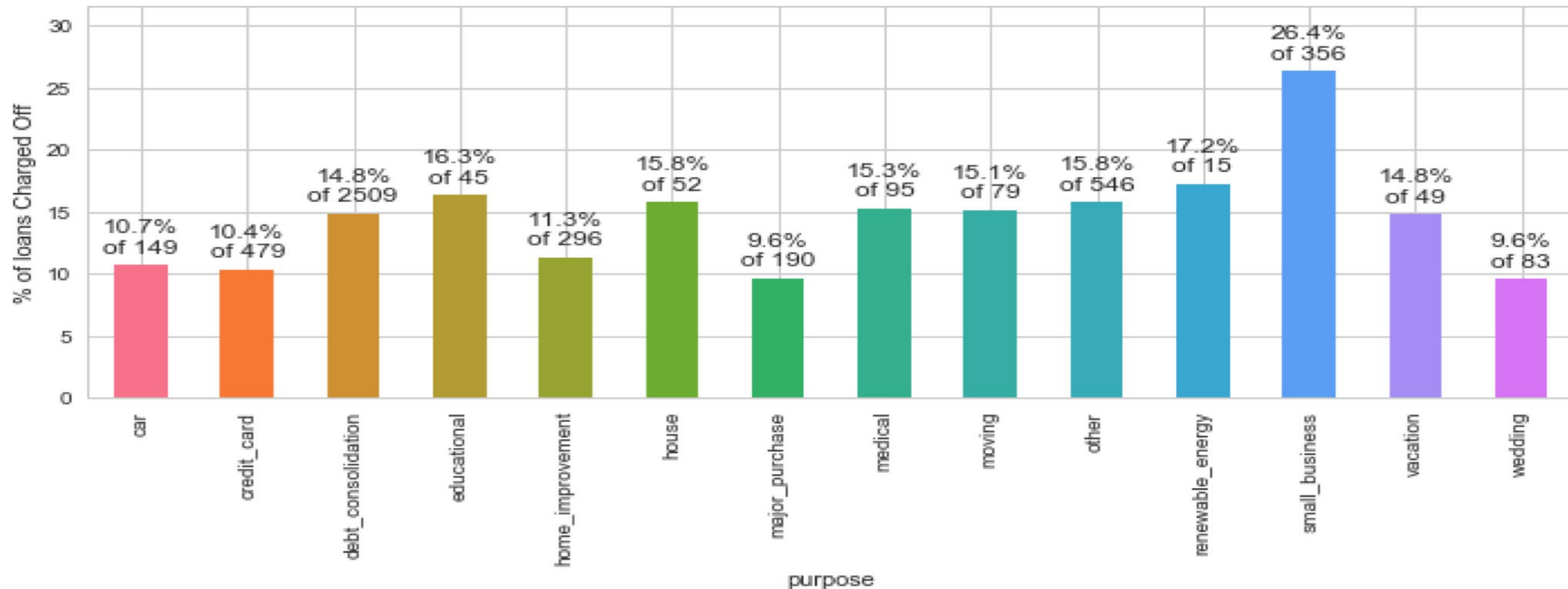


Analysis - Defaults by Interest Rate



Analysis - Defaults by Loan Purpose

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

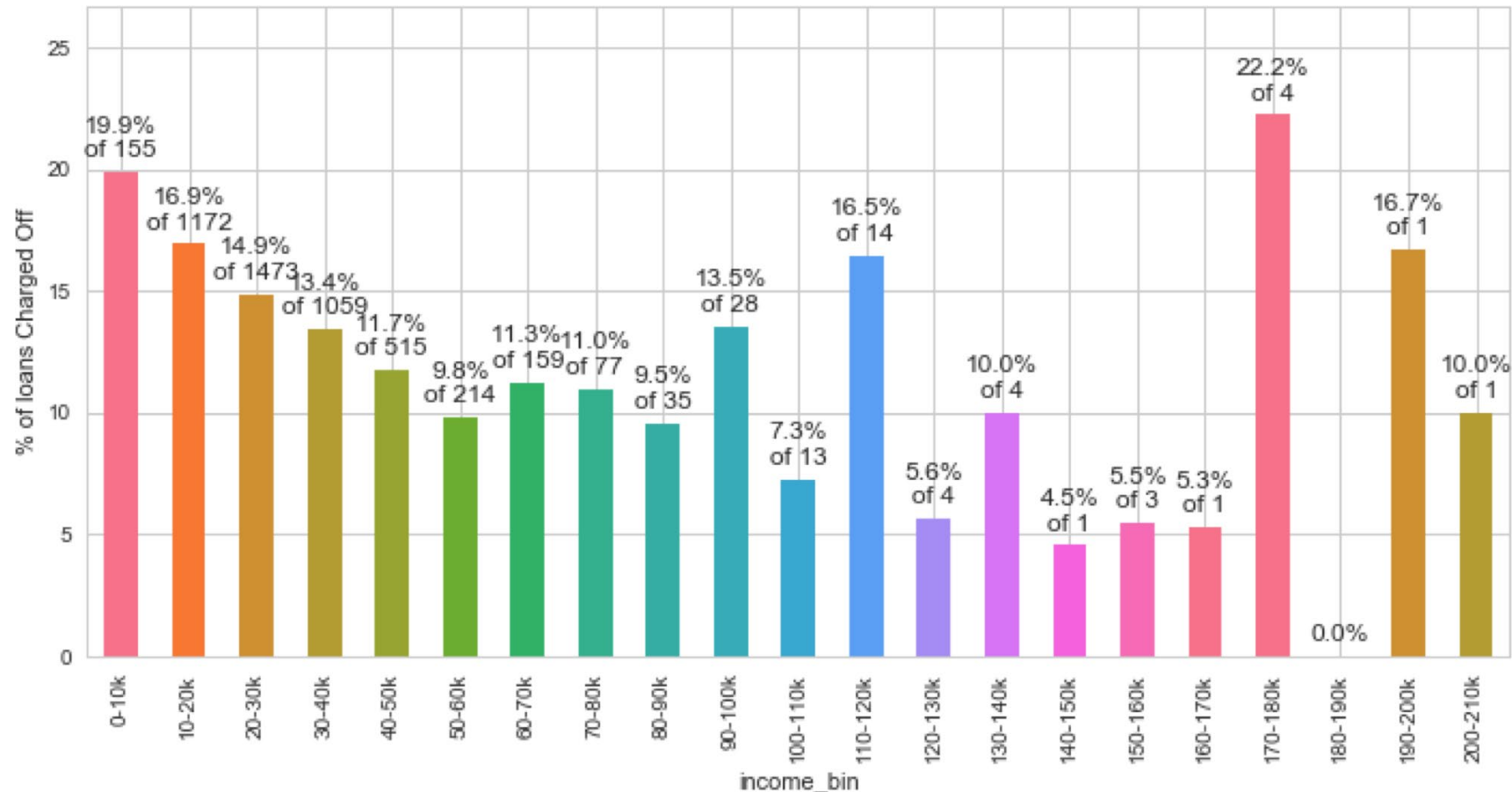


Analysis - Defaults by Borrower's Income

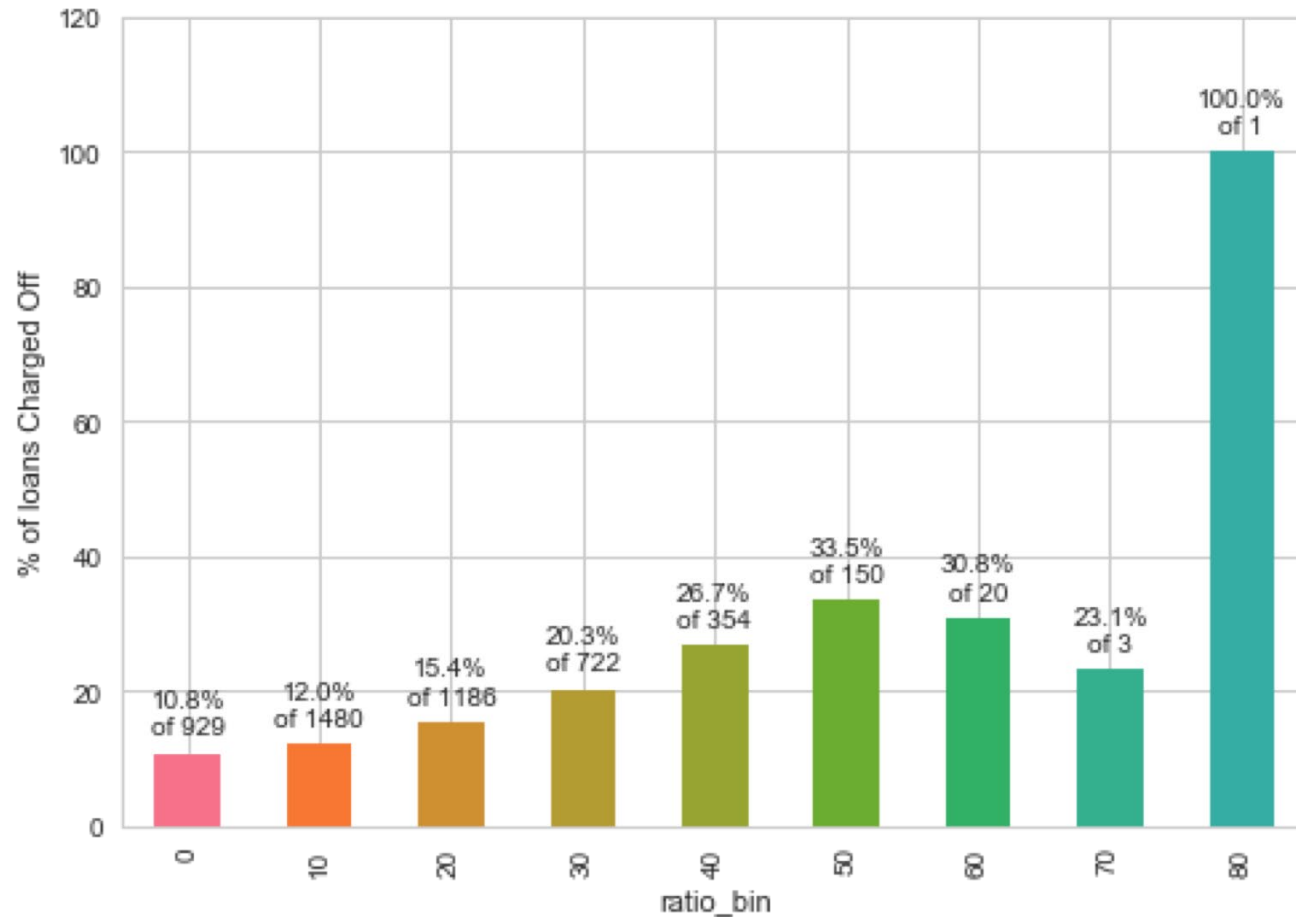
Borrowers having annual income less than 20000 default on their loans at much higher rates.

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

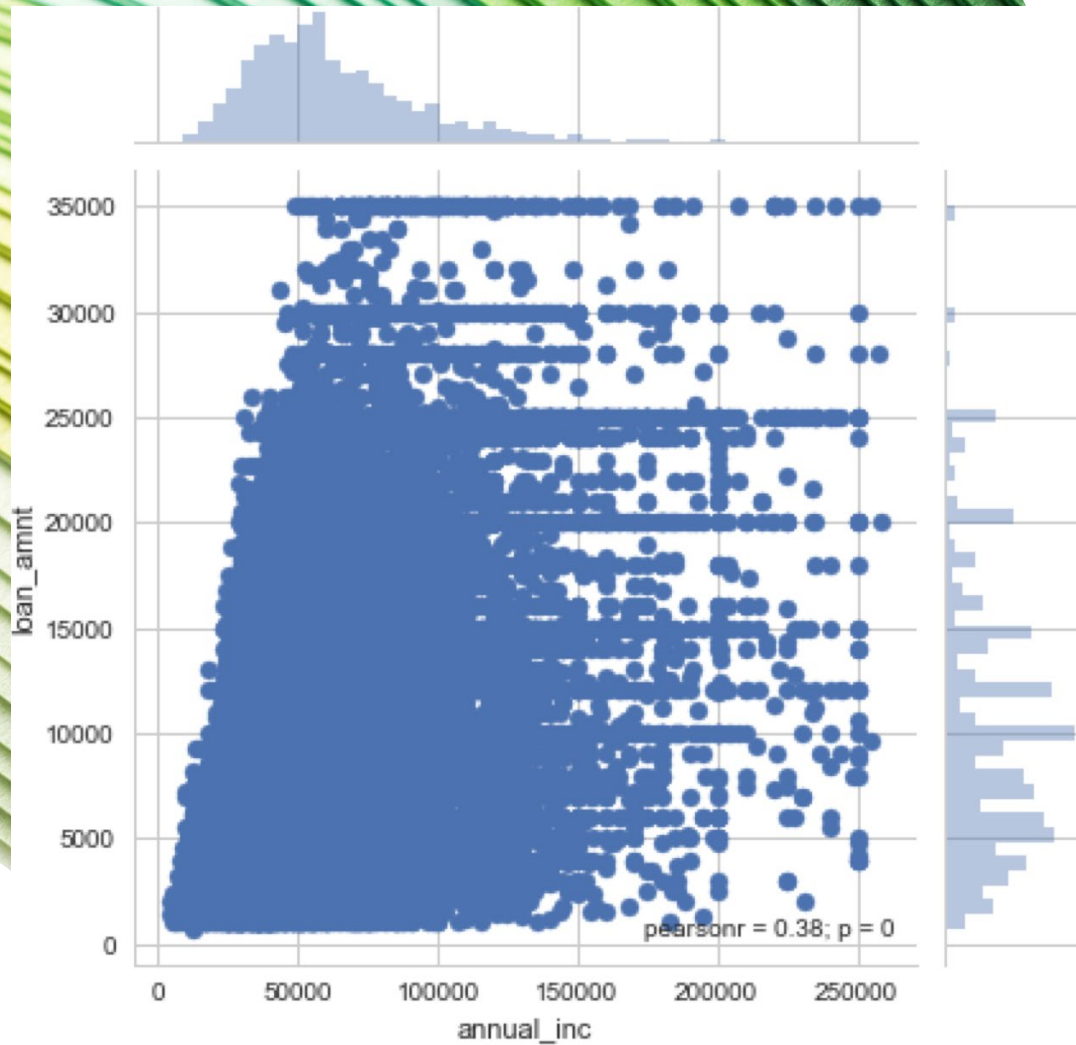


As long as loan amount is less than 20% of annual income, defaults are low.

Loan amounts of 30% of annual income or higher see a high rate of default.

X-axis is the % of loan amount versus annual income

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.

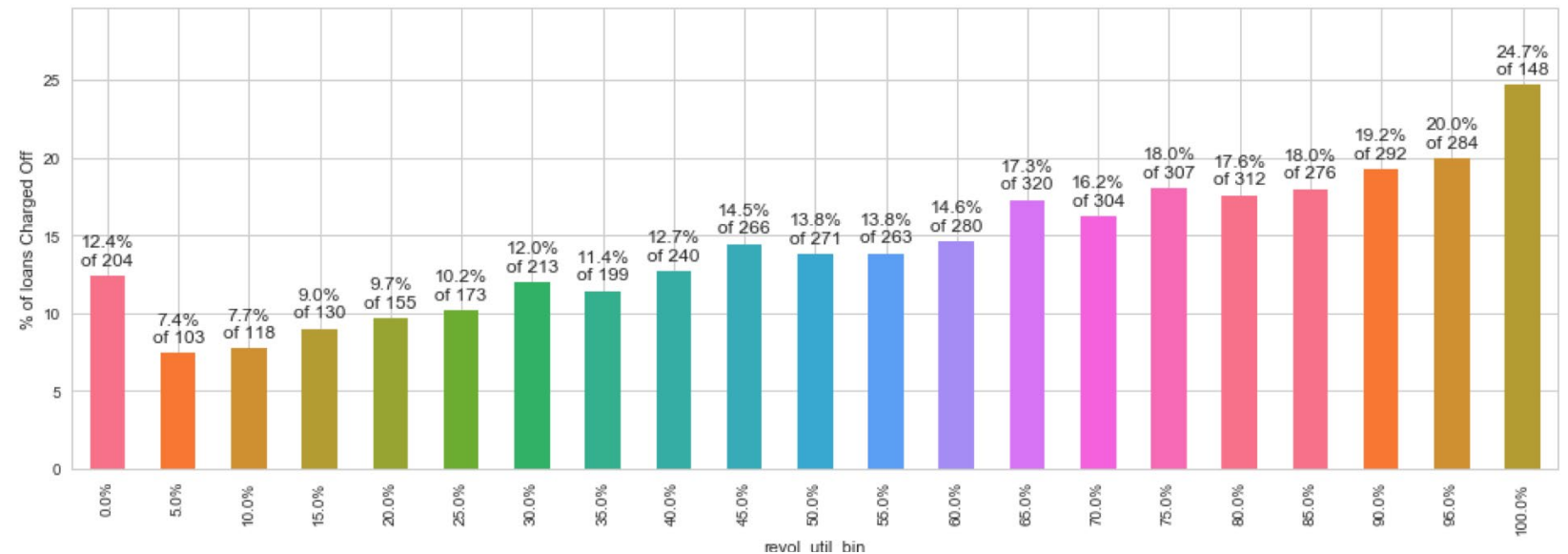
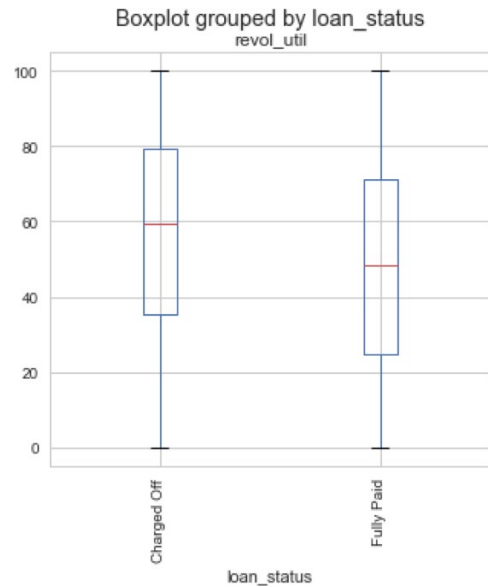


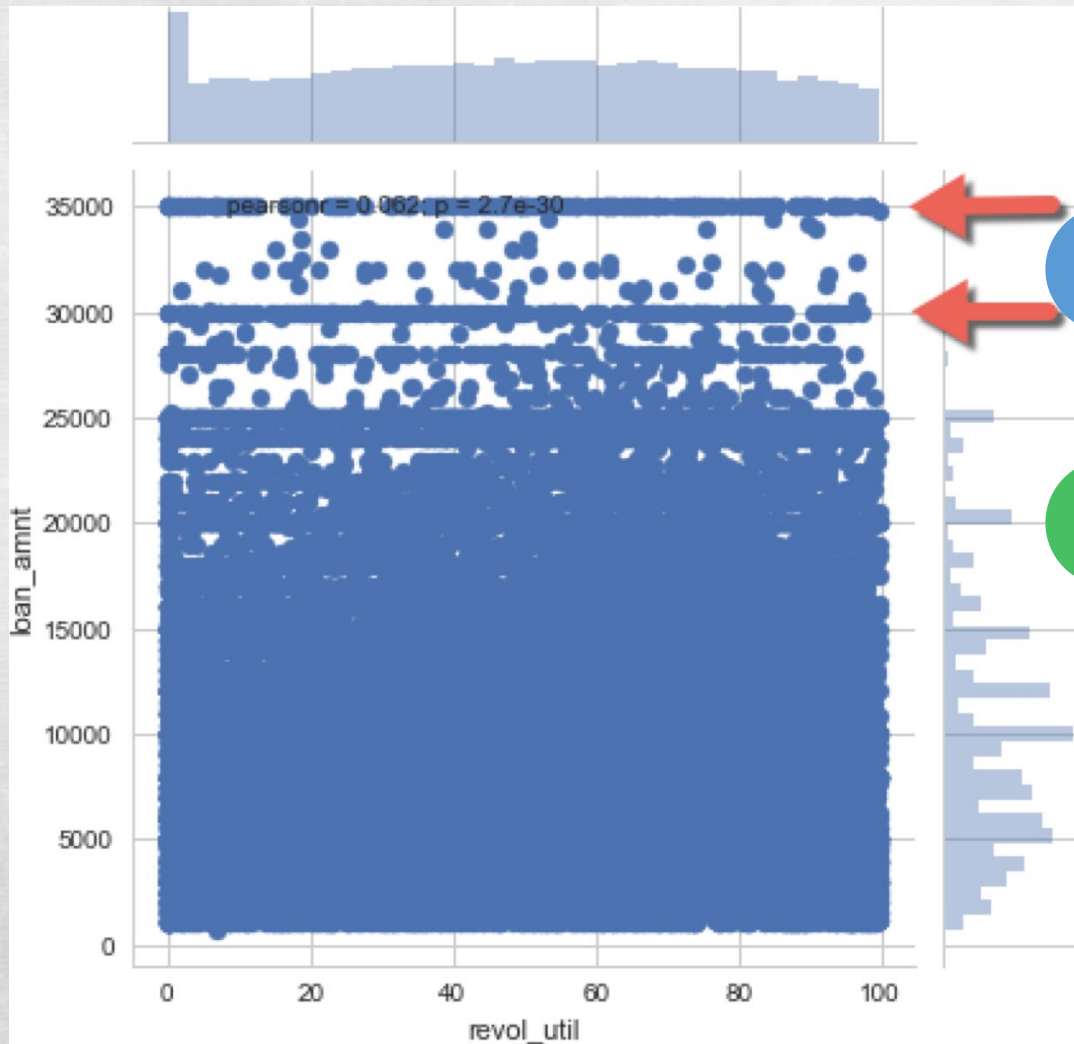
THIS PRACTICE
SHOULD BE
CURTAILED.

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.





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



This practice should be stopped.



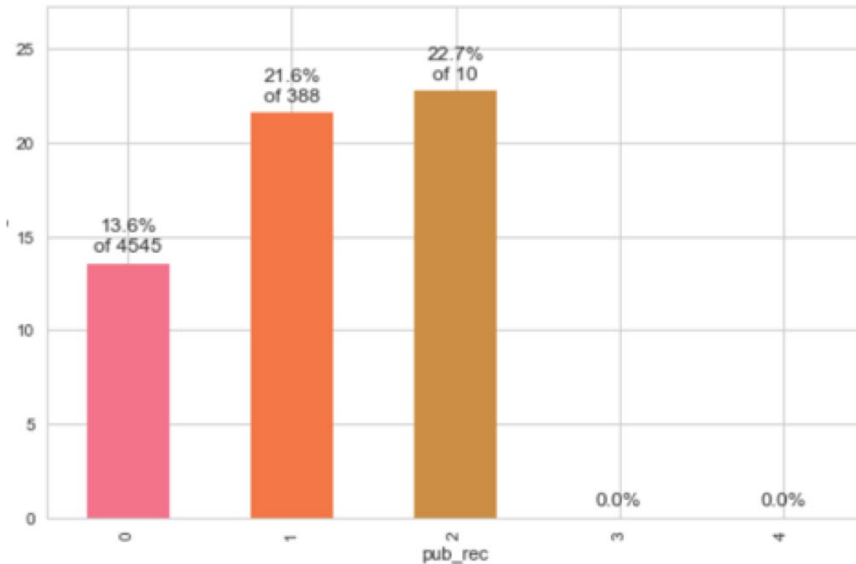
This is risky



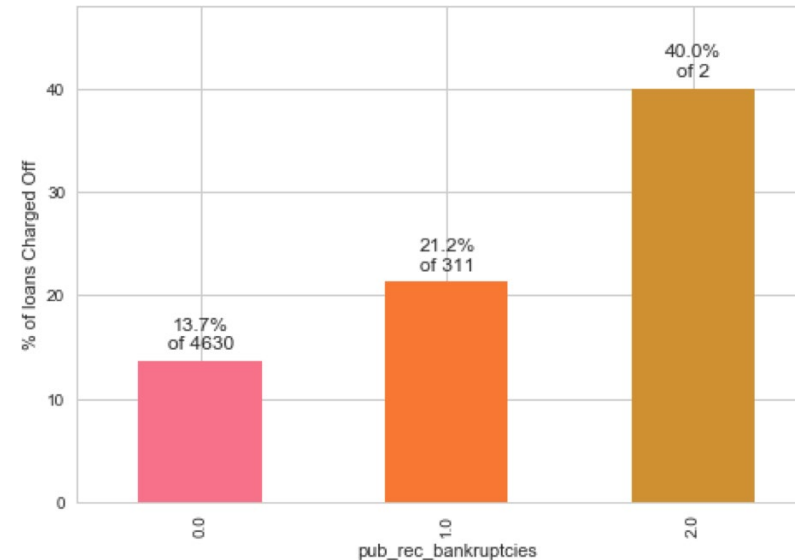
Density of low value loans is also high. They should be approved less often.

Analysis - Defaults by prior bad record

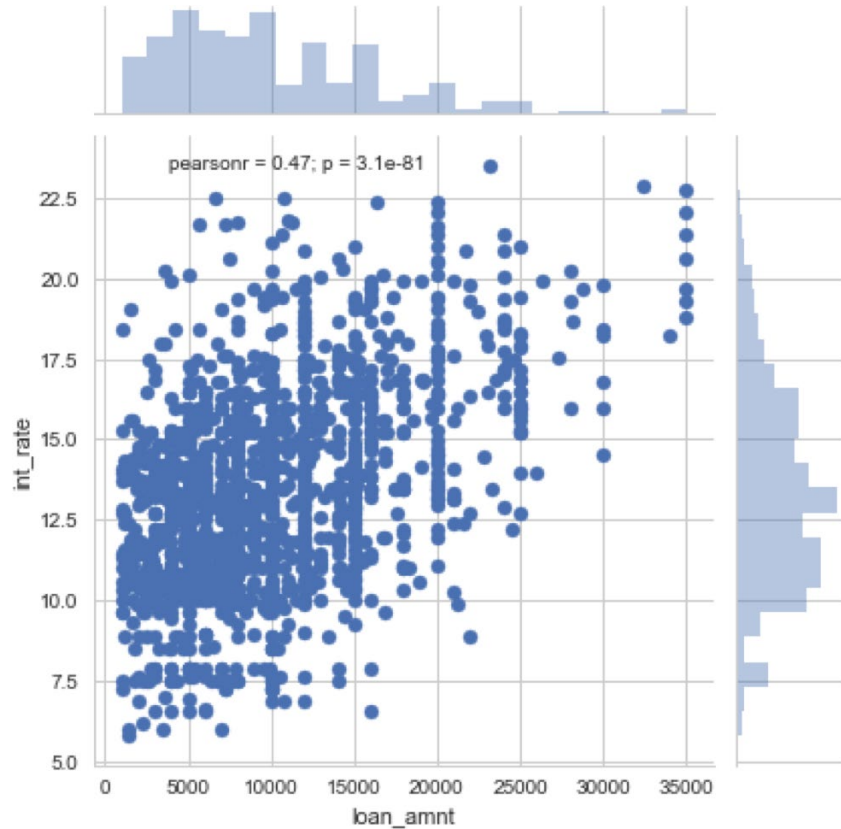
- 94% have no Public derogatory records.
- Having even 1 derogatory record increases the chances of Charge Off significantly.



- 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

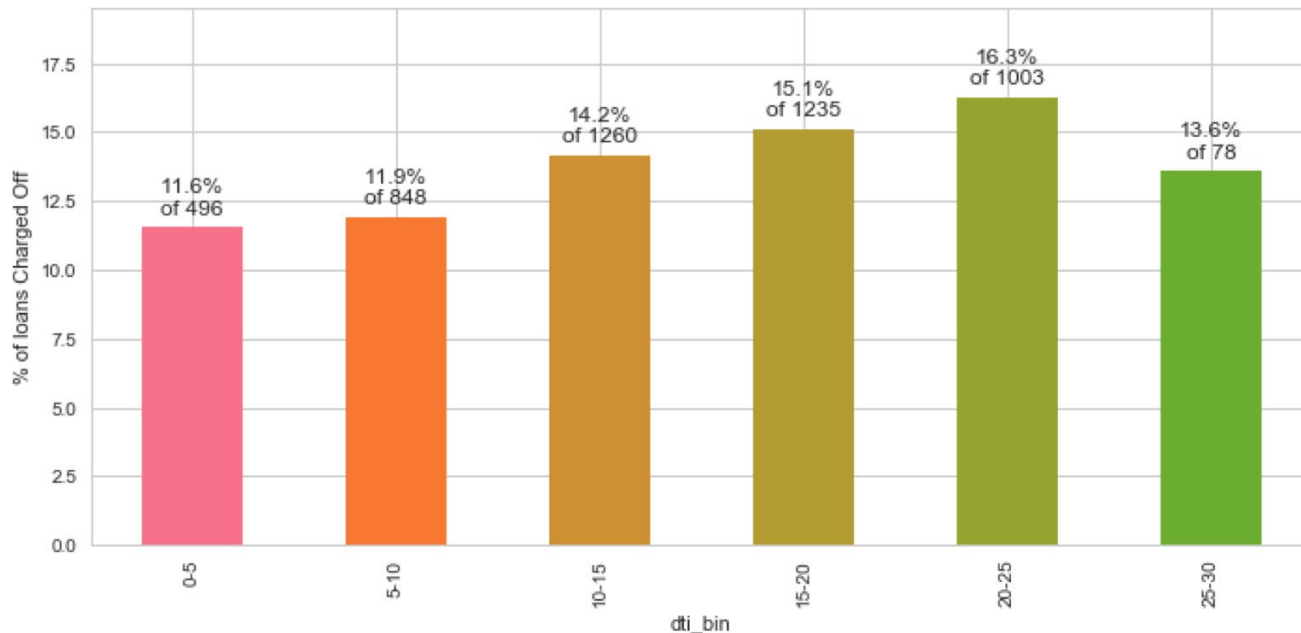
High value loans, as well as low interest loans have been extended to those with prior public derogatory records.

This practice can be stopped to improve business metrics.

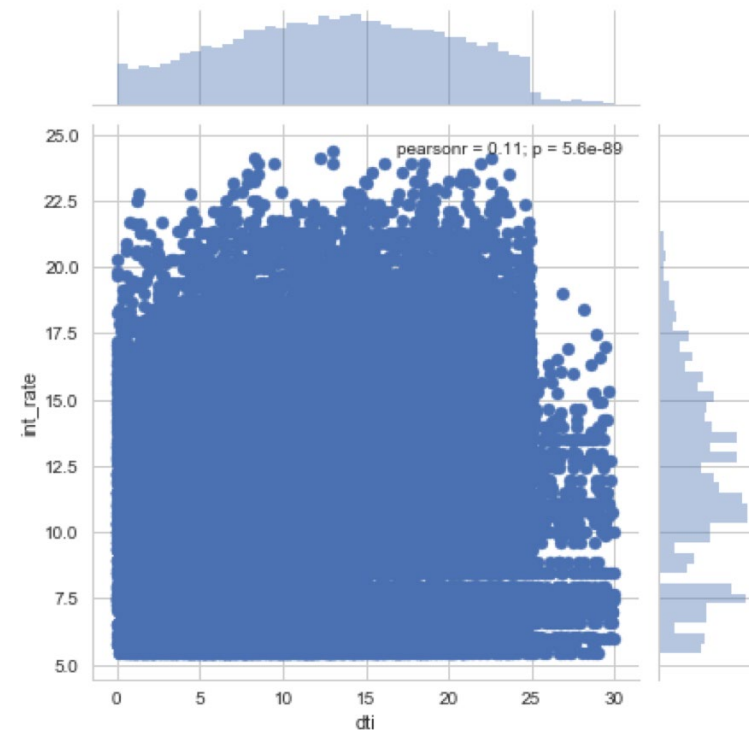
Analysis – Defaults by Debt to Income Ratio

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





Suggestions – for better borrowers

- Stop approving loans where amount/income is higher than 30%
- Reduce approvals for borrowers with purpose as small business
- Stop approving high value loans when revolving line utilization rate is greater than 75%
- Stop approving loans to people with prior bad record
- Stop approving very high value loans
- Start charging higher interest rates for loans with dti greater than 20



End