Presentation

Lending Club Case Study

Presentation to share the overall approach to the analysis done on the data to understand driving factors behind Loan Default

Lending_Club_Case_Study (github.com)	09TH February '22

Anjali Vashisth *Vishnu Vardhan Reddy*

UPGRAD

PG Programme In Machine Learning and AI

Problem

CONTEXT

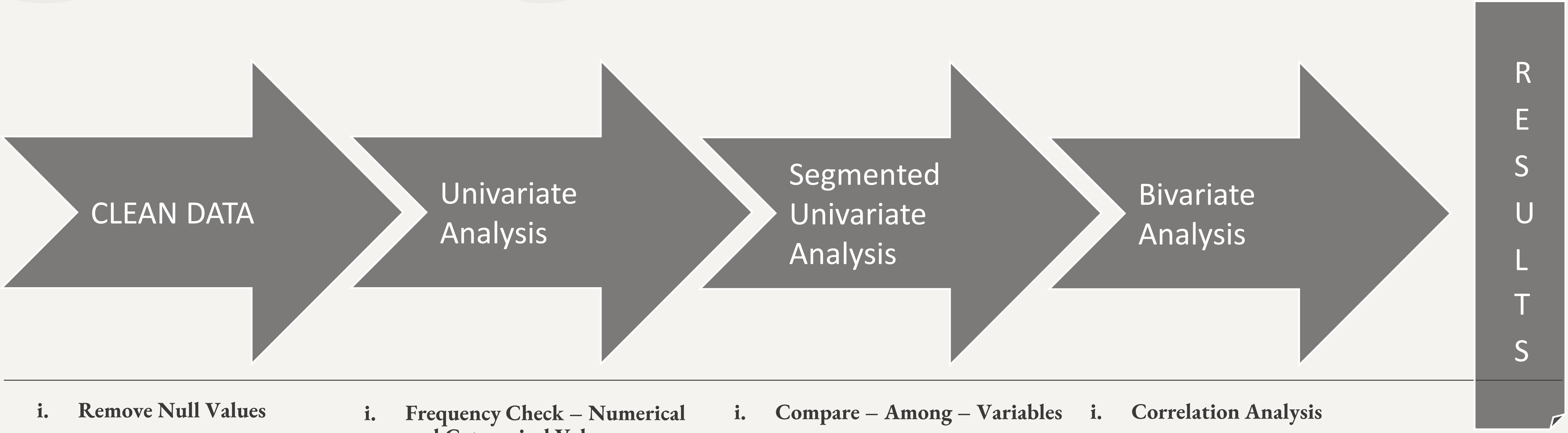
Lending Club wants to understand the driving factors behind loan default, i.e. the driver variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.

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.

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



i. **Remove Null Values**
ii. **Correct Data Type**

Cleaning

i. **Frequency Check – Numerical and Categorical Values**
ii. **Create – Derived Variables**

Univariate Analysis

i. **Compare – Among – Variables**
ii. **Create Derived Variables**

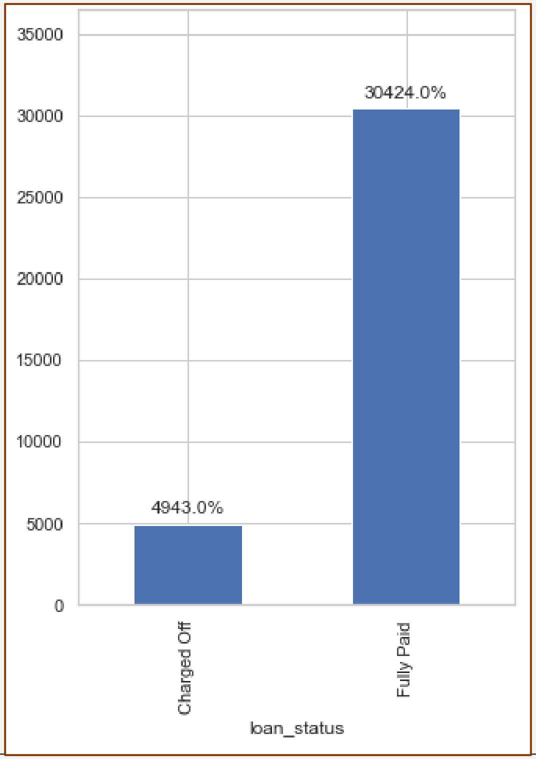
Segmented Bivariate

i. **Correlation Analysis**
ii. **Analyse – Joint Distributions**

Bivariate Analysis

Publish observations

Analysis - Overall Loan Status

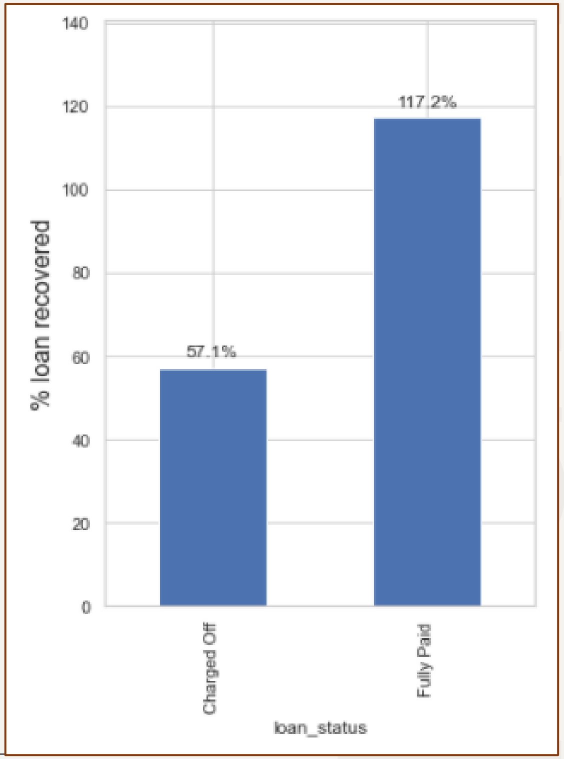


TOTAL LOANS

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

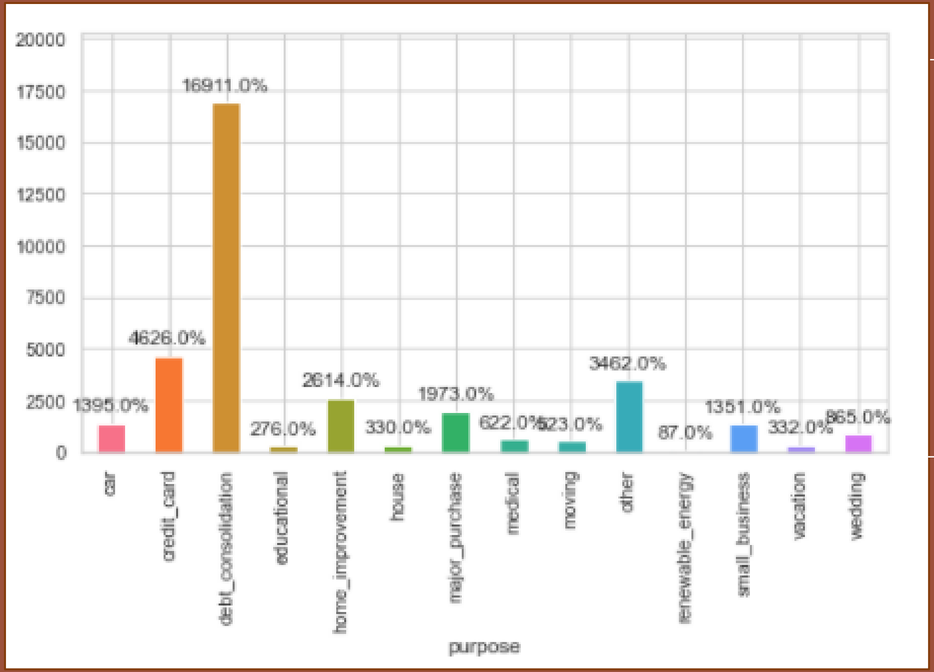
Total Money Earned

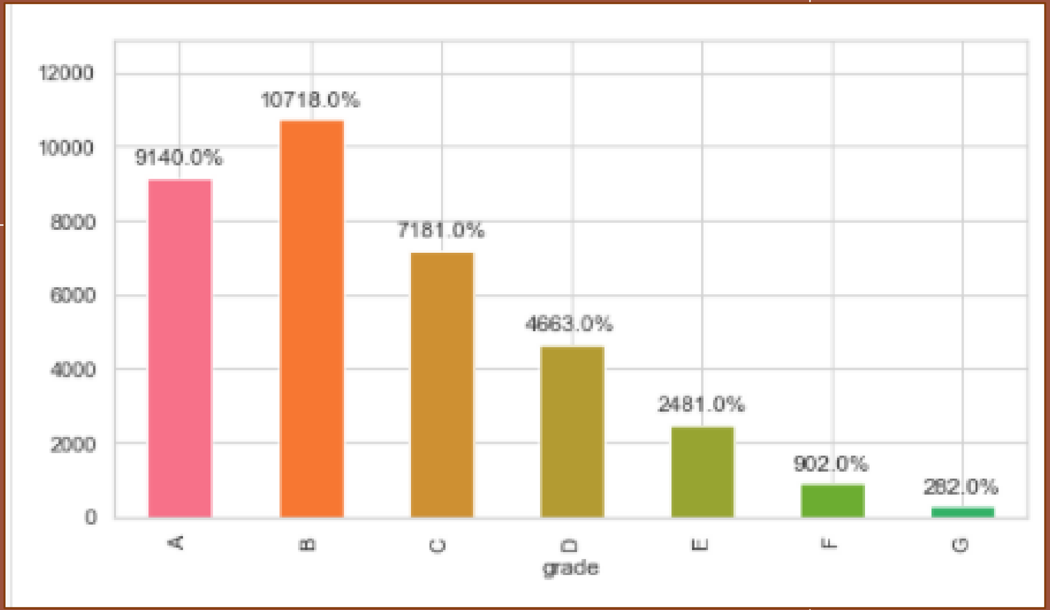
*Lending Club only **recovers 57%** of the loan amount when loans are defaulted. On fully paid up loans, the company makes 17% profit.*



Analysis- Understanding Loans

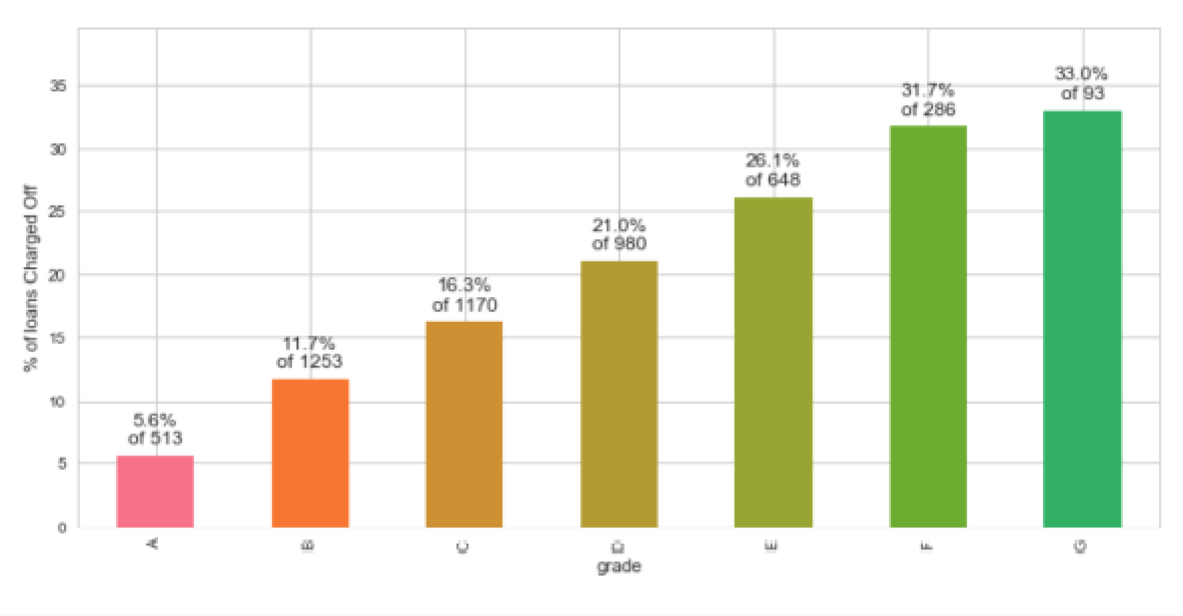
Maximum number of loans are for debt consolidation, followed by credit card





Grade A and B have mostly high Loans

Analysis - Understanding Loans Continued

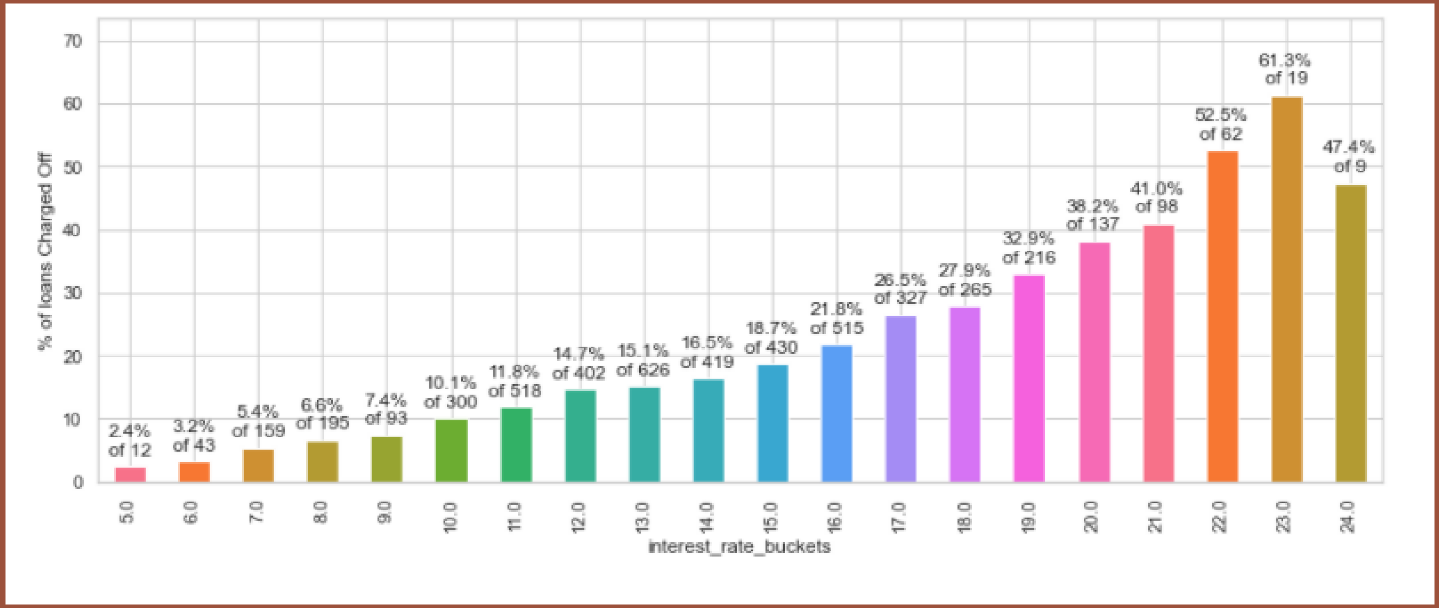


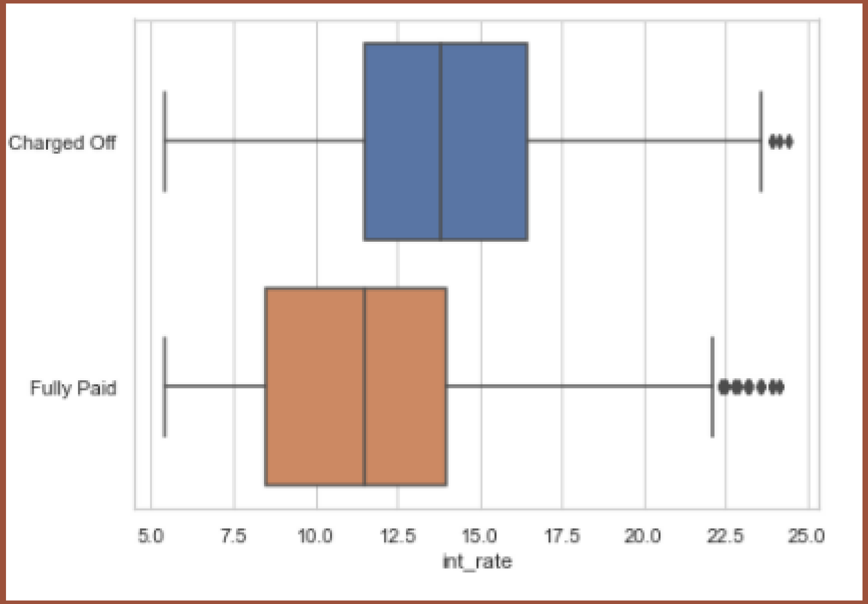
Lower grades have higher incidence of defaults on loans. The grading system is working!

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

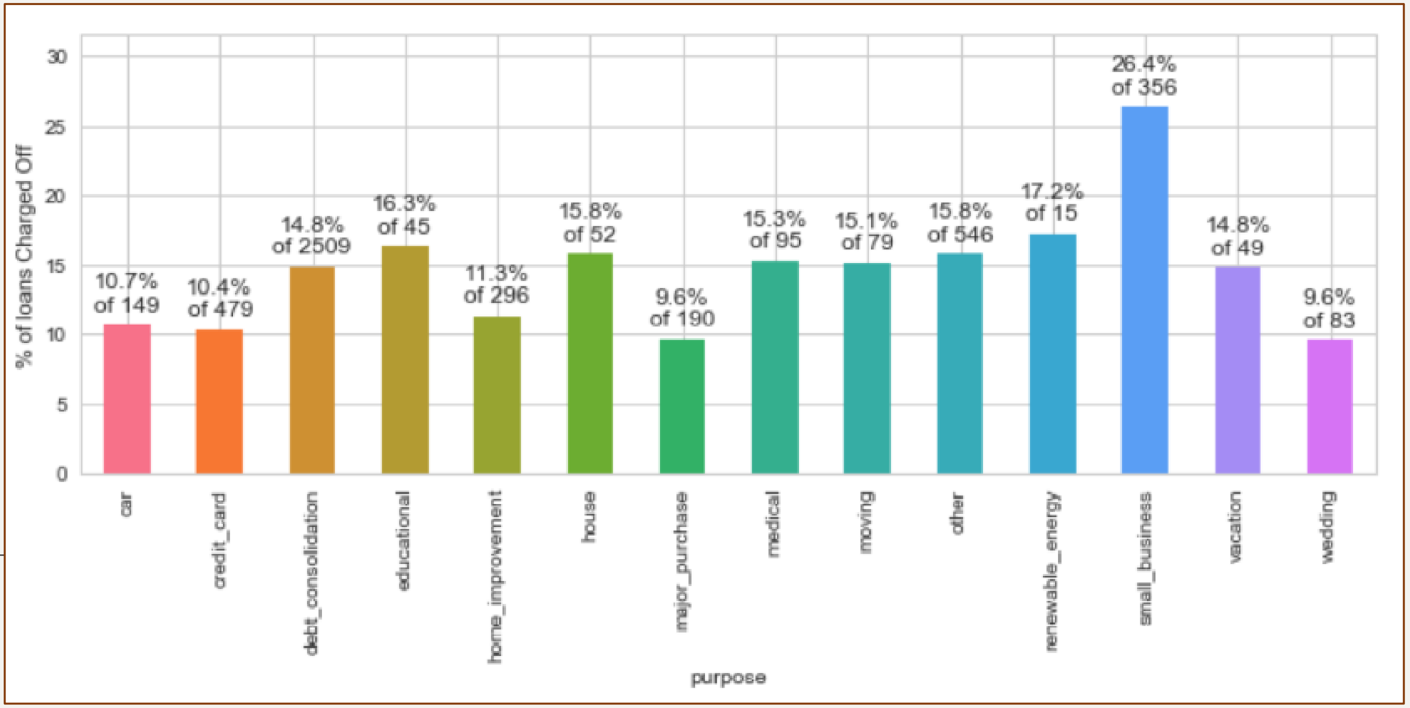
Percentage of Defaults increases monotonically with higher interest rates. At rates of 19% and above, more than 33% of loans are Charged Off.



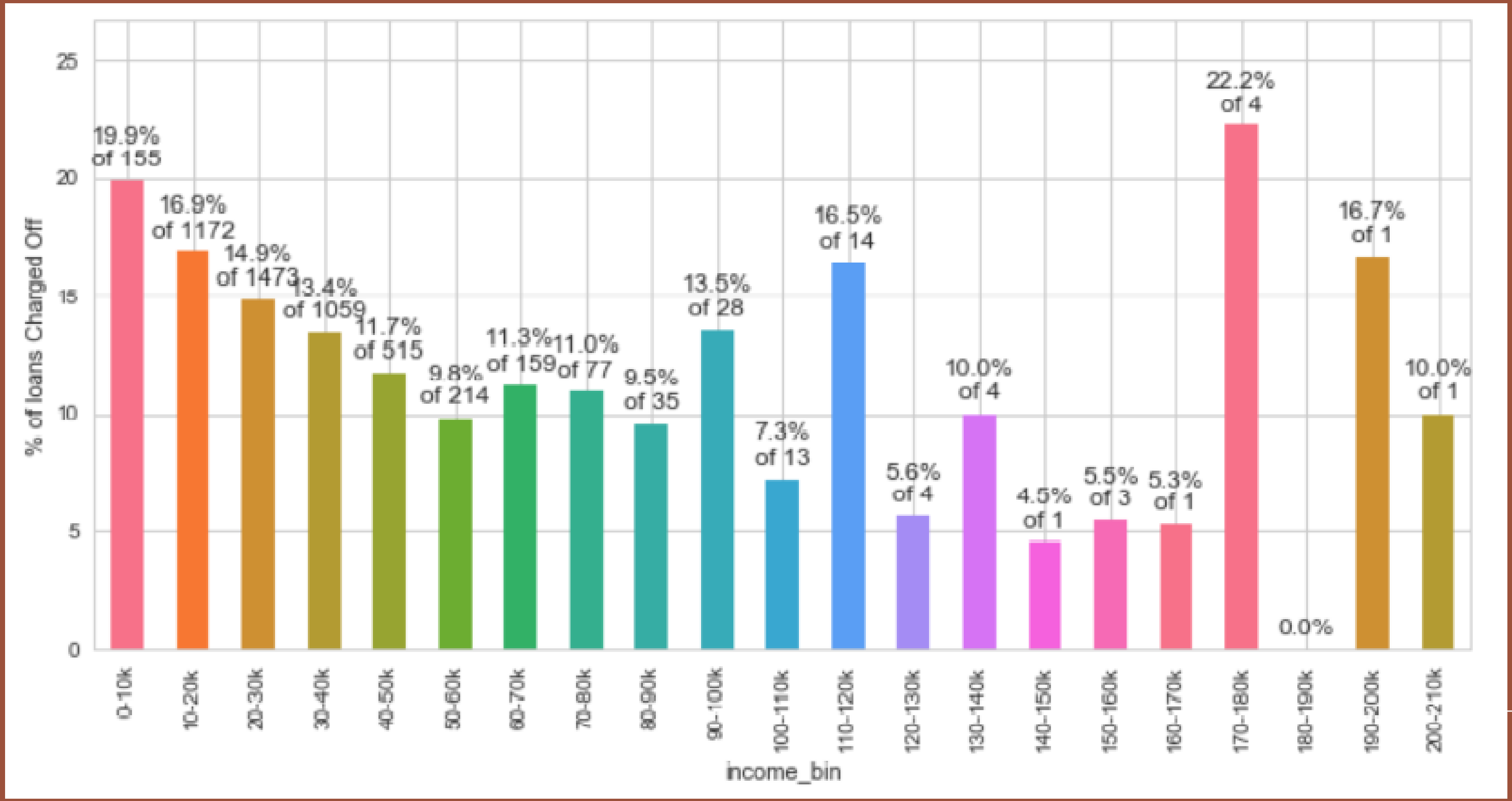


Analysis - Defaults by Loan Purpose

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



Analysis - Defaults by Borrower Income



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

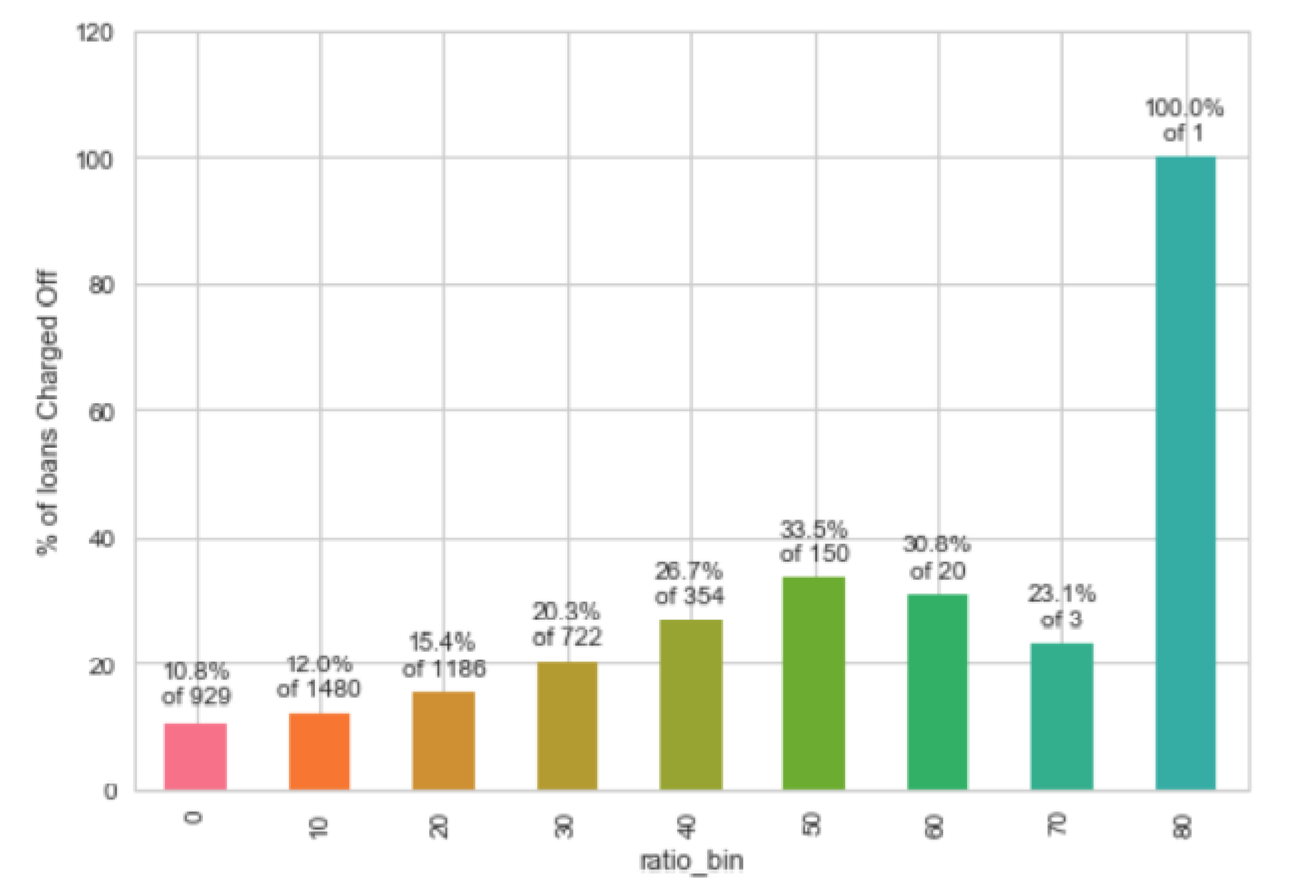
As we will see on next slide – the ratio of amount to income is more important.

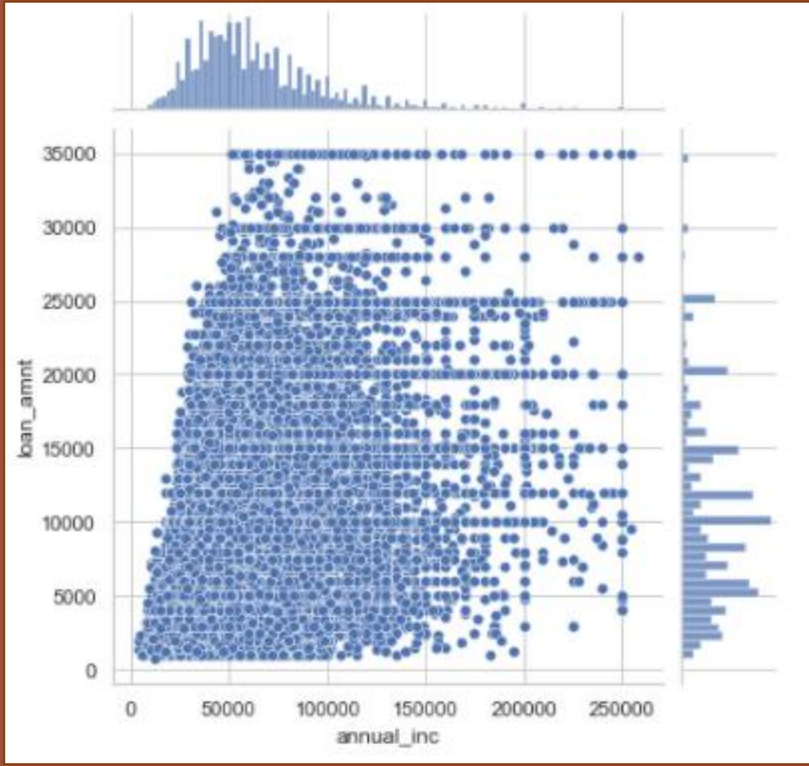
Analysis - Defaults by ratio of amount to income

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

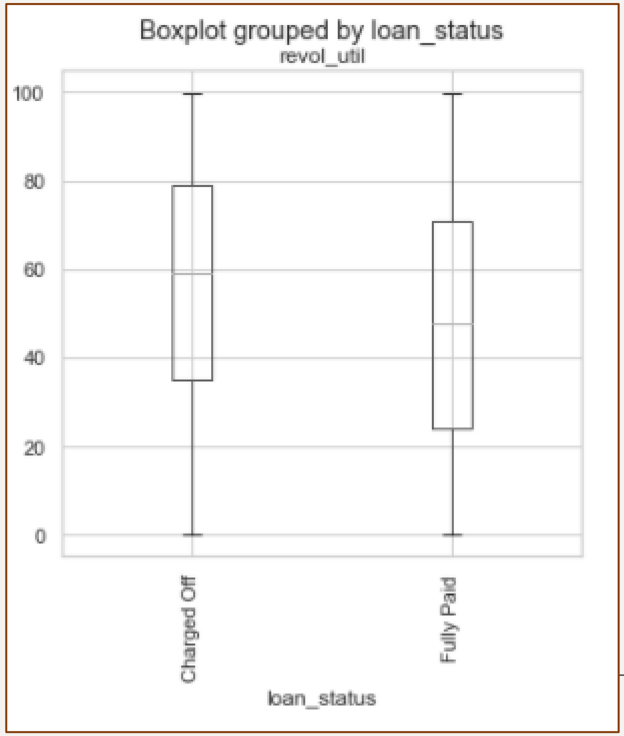




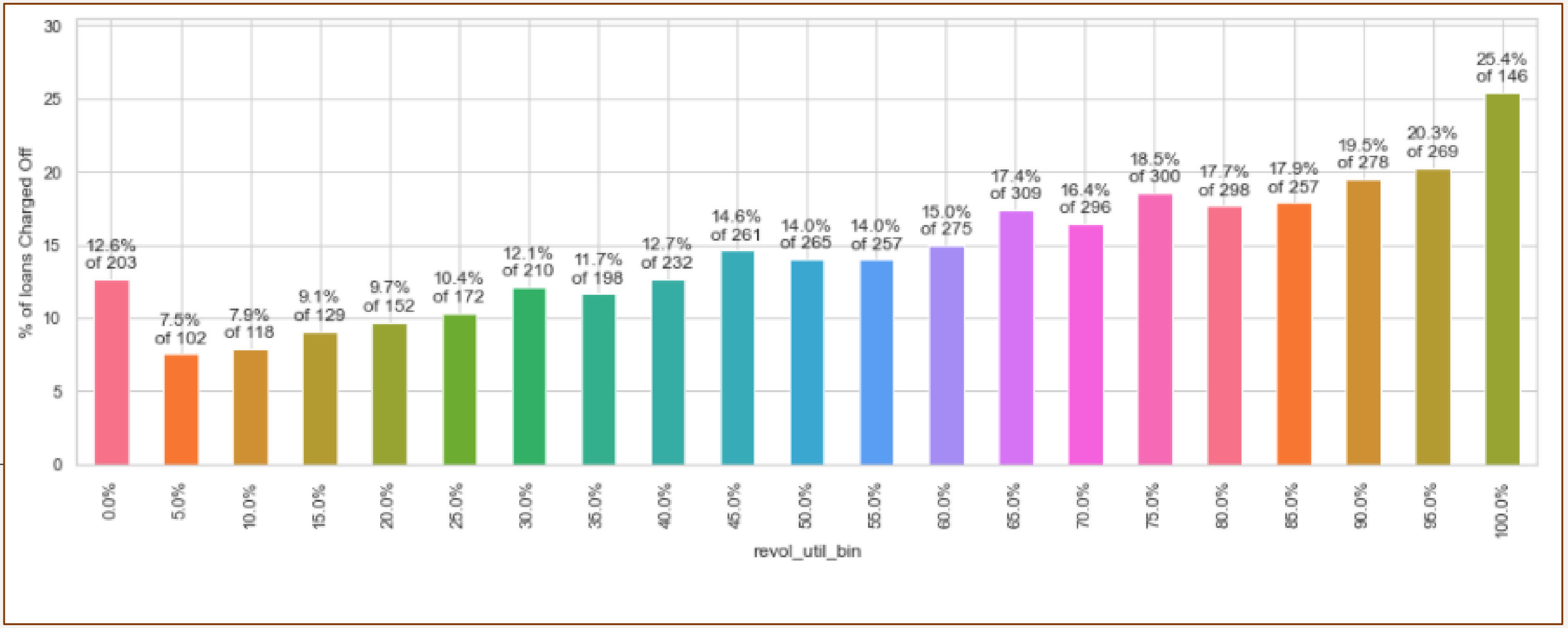
- We see here that Lending Club has extended high-value loans to people with low income.
- There are many cases of people with income 50000 or less getting loans of 25000 or more.
- This practice should be curtailed.

Analysis - Defaults by ratio Continued

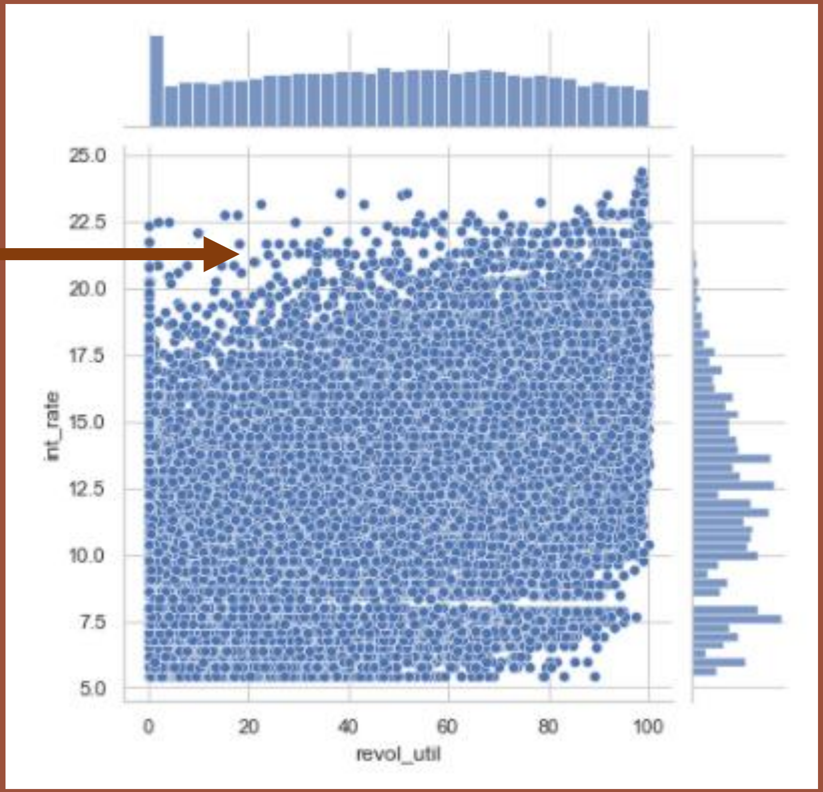
Analysis - Defaults by Revolving Line Util Rate



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



Analysis - Defaults by Rev Util Continued

Analysis - Defaults by prior bad record

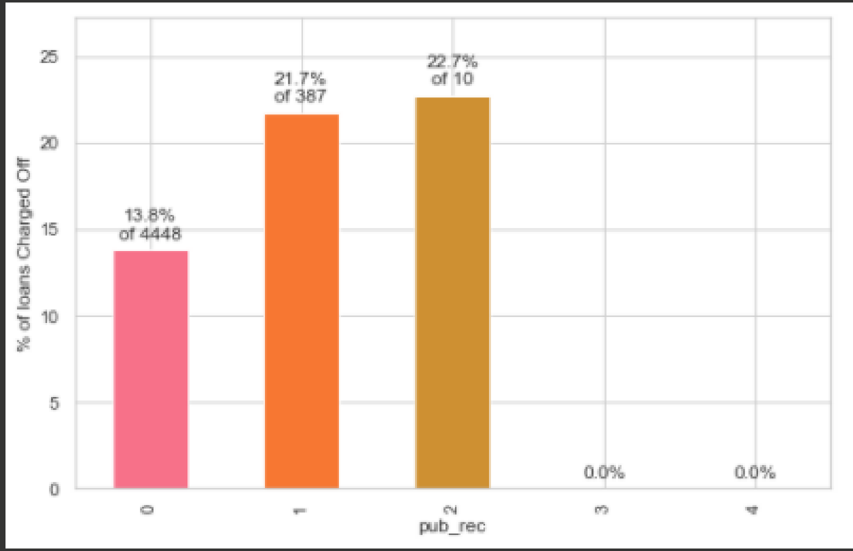
96% have no bankruptcy record.

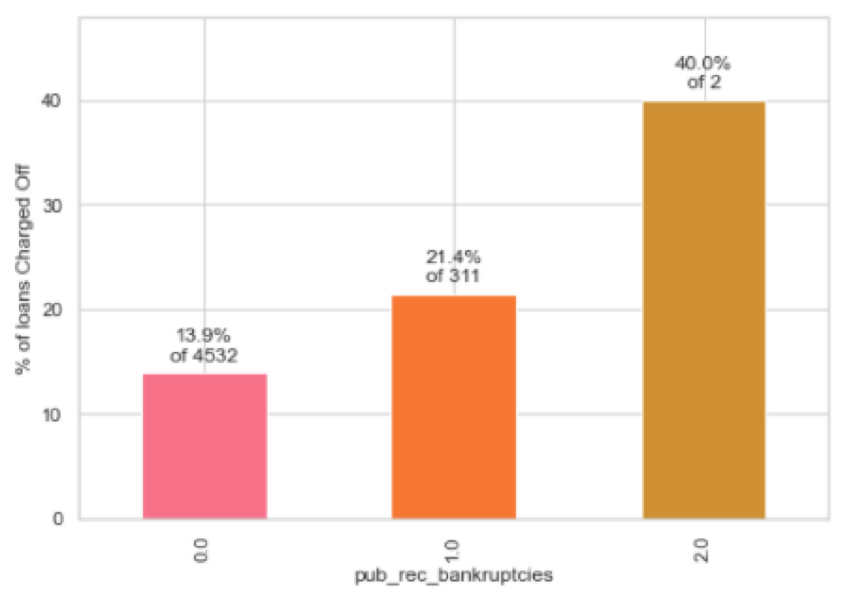
Having even 1 bankruptcy record increases the chances of Charge Off significantly.

Public Derogatory Record and Public Bankruptcy records have 83% correlation. We can use any one.

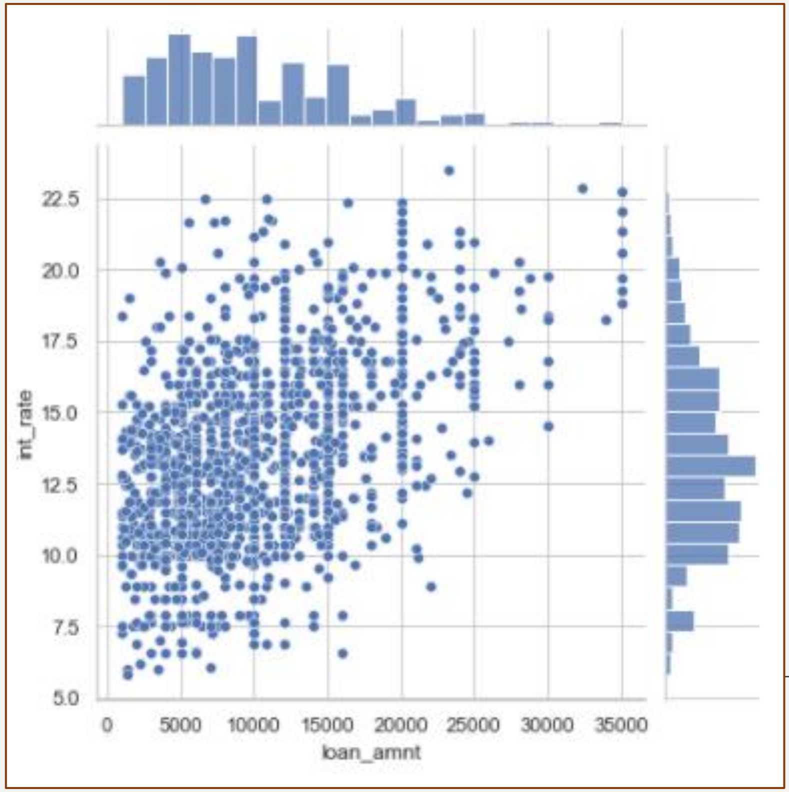
94% have no Public derogatory records.

Having even 1 derogatory record increases the chances of Charge Off significantly.





Analysis by prior bad record - Continued

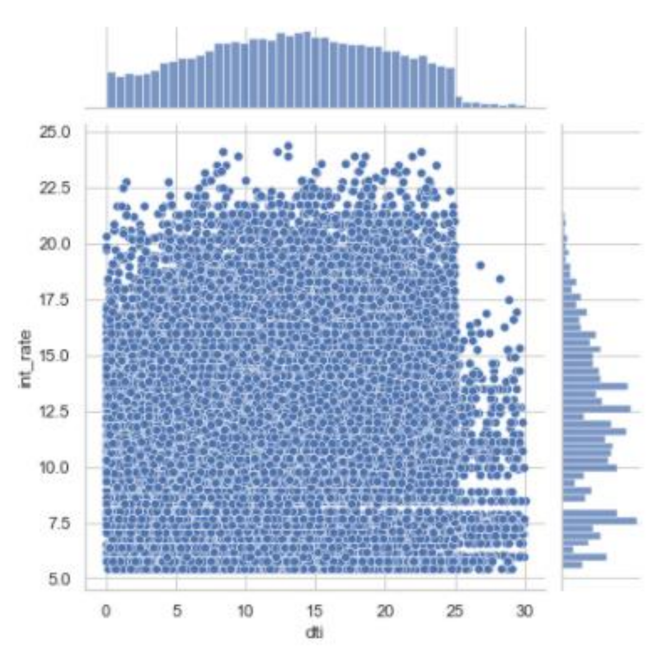


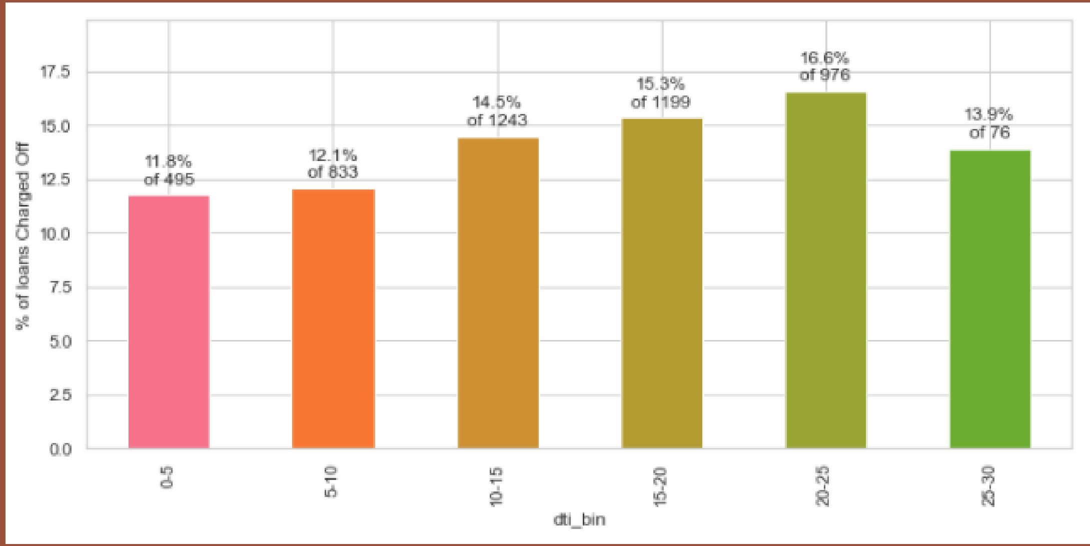
High value loans, as well as low interest loans have been extended to those with prior public derogatory records. This practice can be stopped to improve business metrics.

Data of people with >0 bad records

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.

Recommendations

Better Quality Borrowers

STOP approving loans where amount/income is higher than 30%

REDUCE number of approvals where purpose is small business

STOP approving high-value loans when revolving line utilization rate greater than 75%

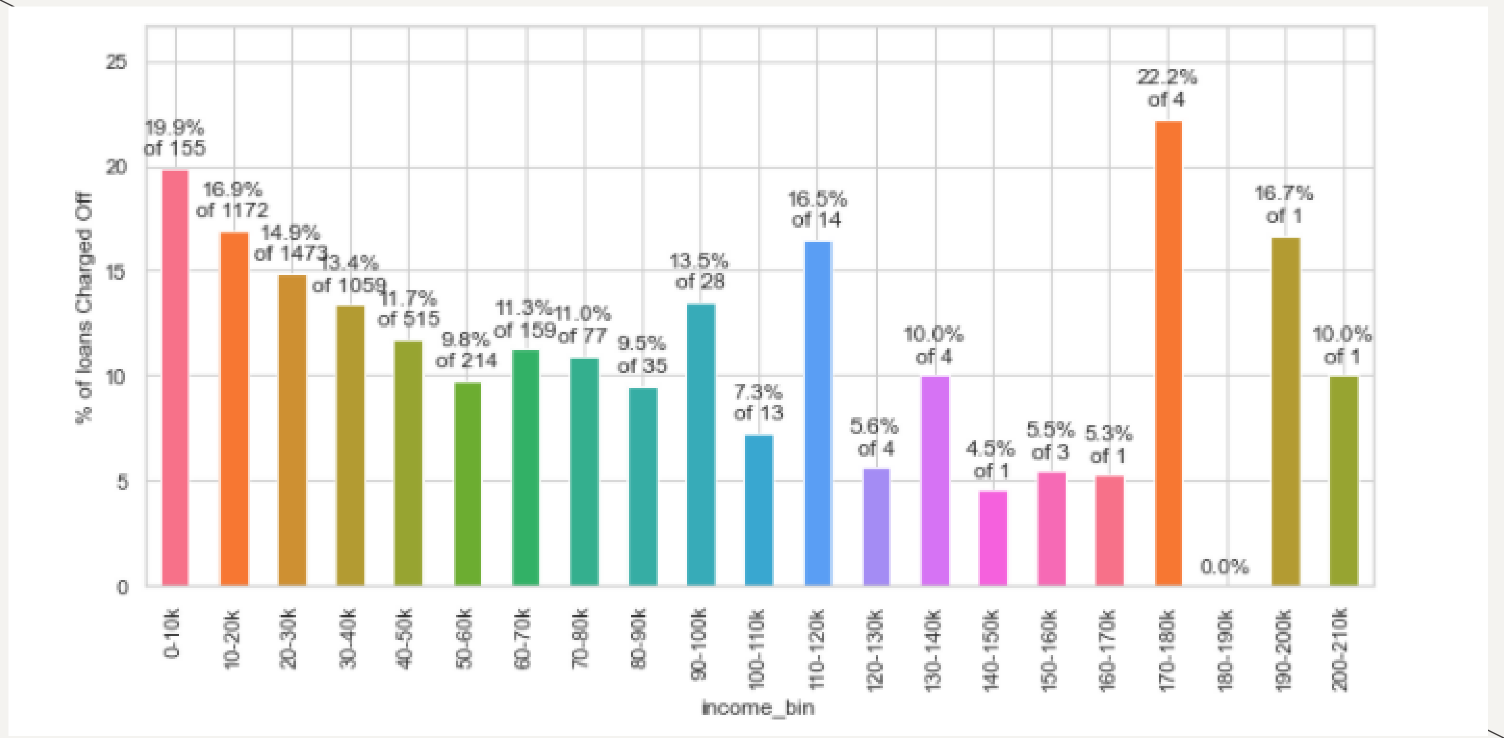
START approving loans to people with prior bad record. Or at least stop approving high-value loans

STOP charging higher interest rates for loans with dti greater than 20

Upgrad

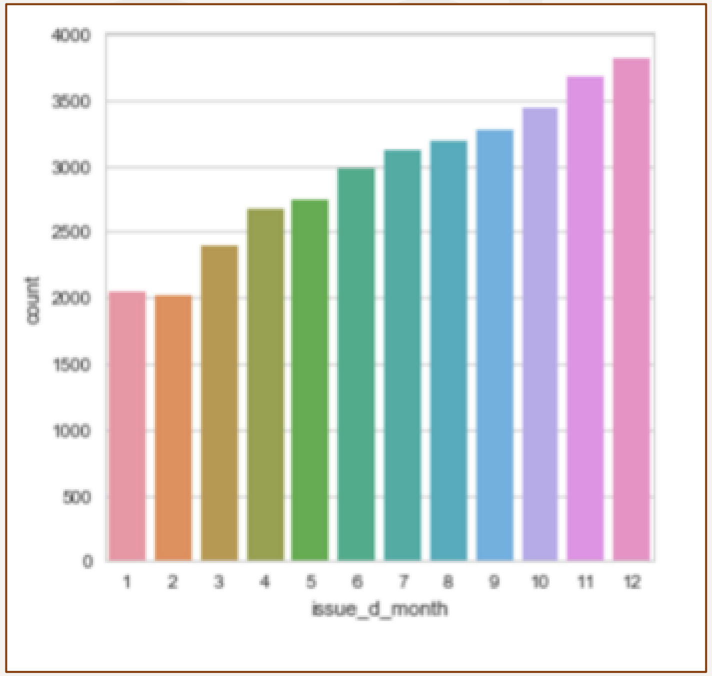
Interesting Findings:

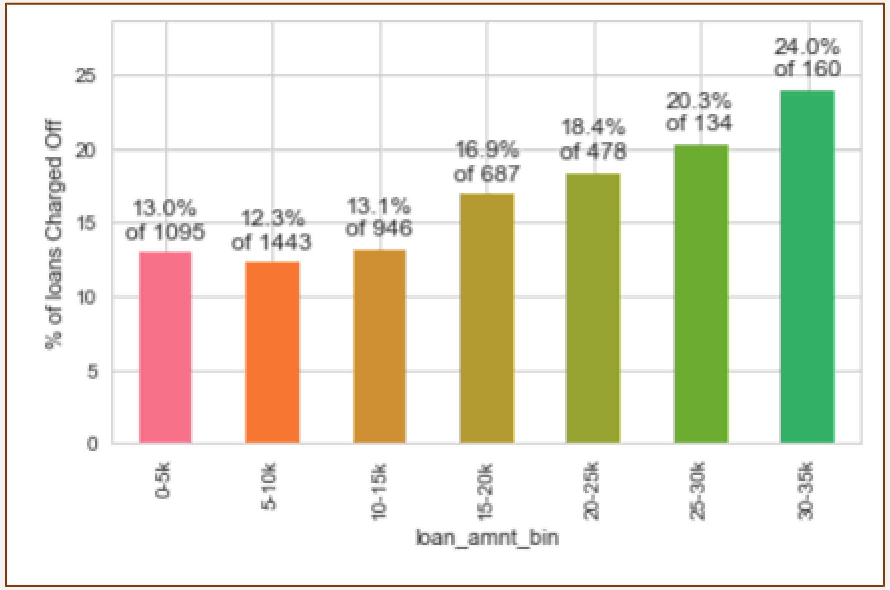
People who took their first loans just before an economic crisis, like the one in 1980 and the subprime crisis in 2008, have higher rates of default. Presumably, these are young people who were affected by the economic conditions of their early career



Interesting Findings:

Loan numbers increase as months get closer to year-end. Are people trying to meet targets?

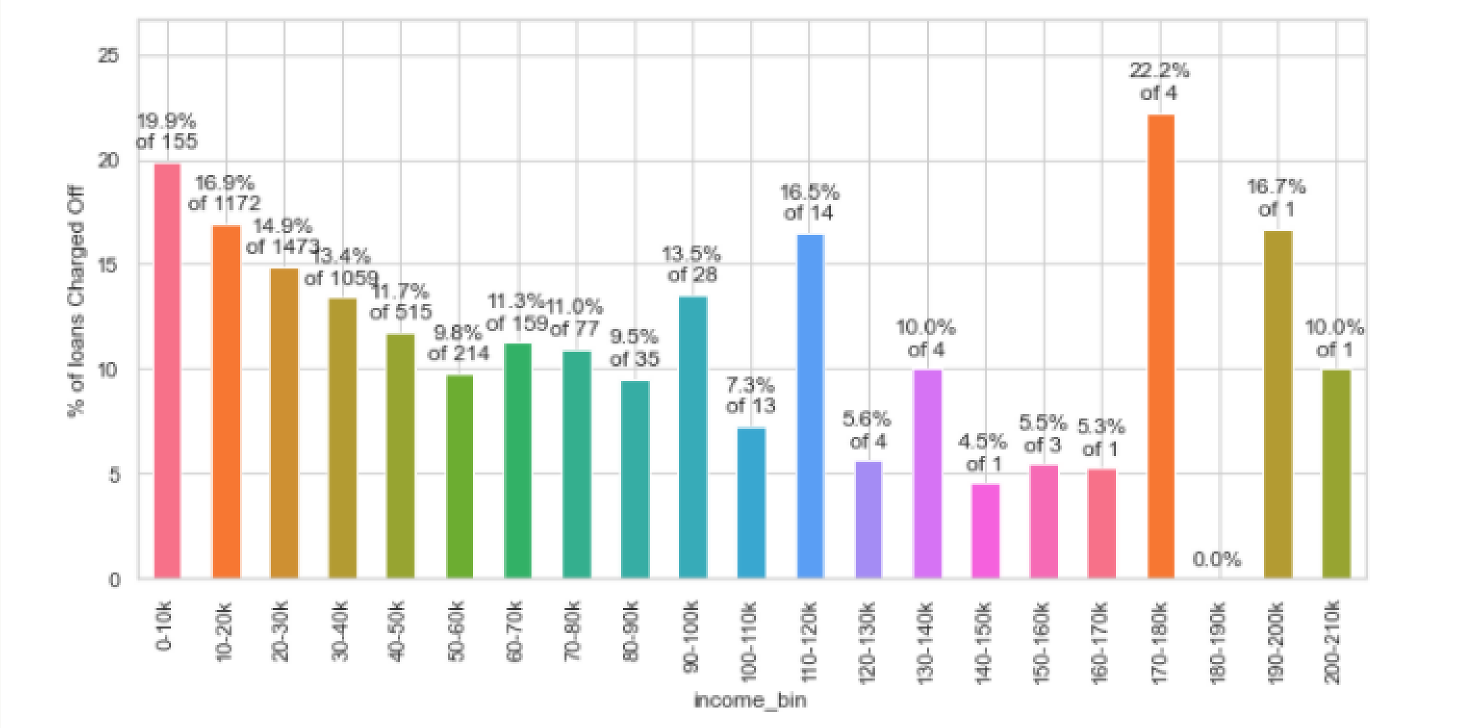




As the loan amount increases, the chance of default also increases.

Interesting Findings:

Maximum loans are in populous states, California, New York, Florida and Texas.
Is it the risk-taking nature of them?



THANK YOU