Lending Club Case Study

SUBMISSION

Group Facilitator: Himabindu Srirangam
Group Member: Oorugonda Anusha

The Problem of Lending Club

Company	Context	Problem Statement
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.	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.	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

Data Cleaning

Univariate Analysis

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

Analyze variables against segments of other variables
> Create derived variables

Segmented Analysis

Bivariate Analysis

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

Publish insights and observations

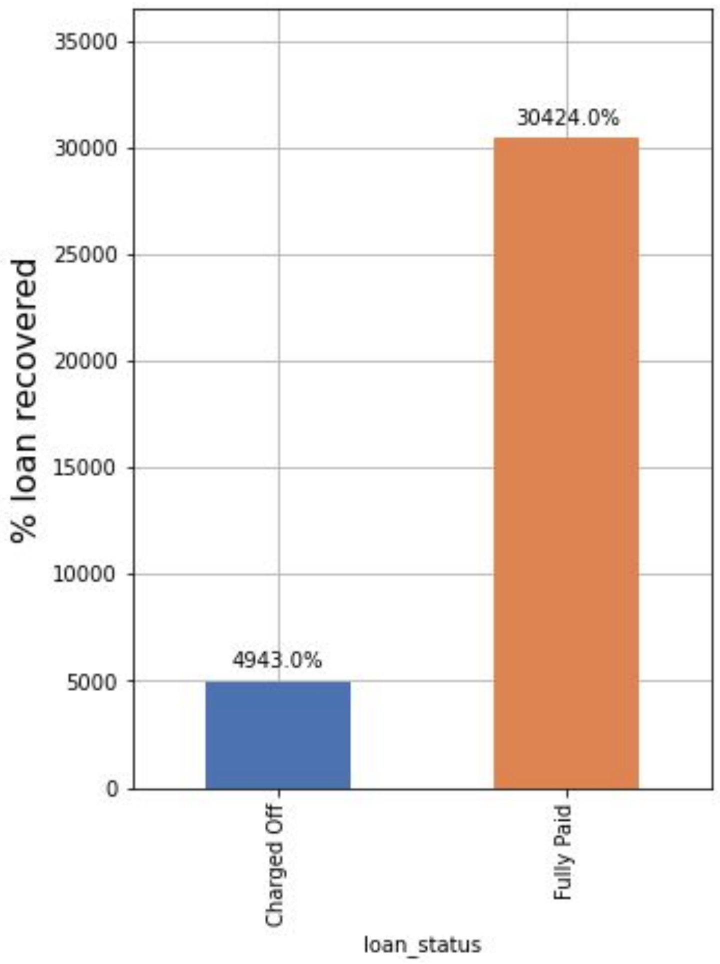
Results & Summary

Analysis - Overall Loan Status

Total Loans

Approx. 14% of the loans are defaulted.

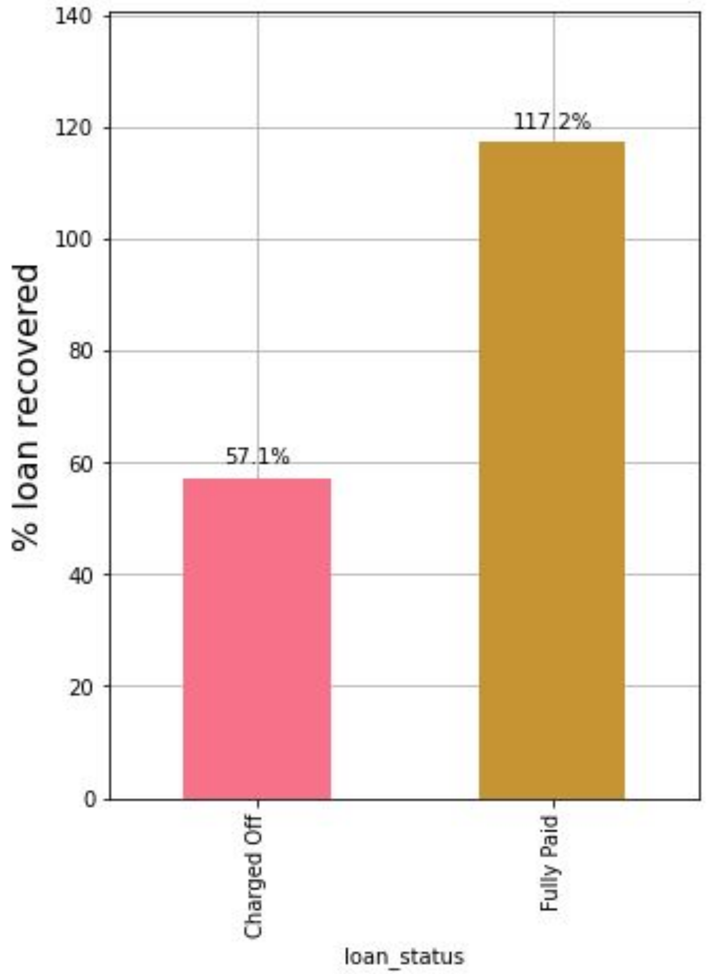
Any variable that increases, % of default to higher than 16.5% should be considered a business risk.



Total money earned

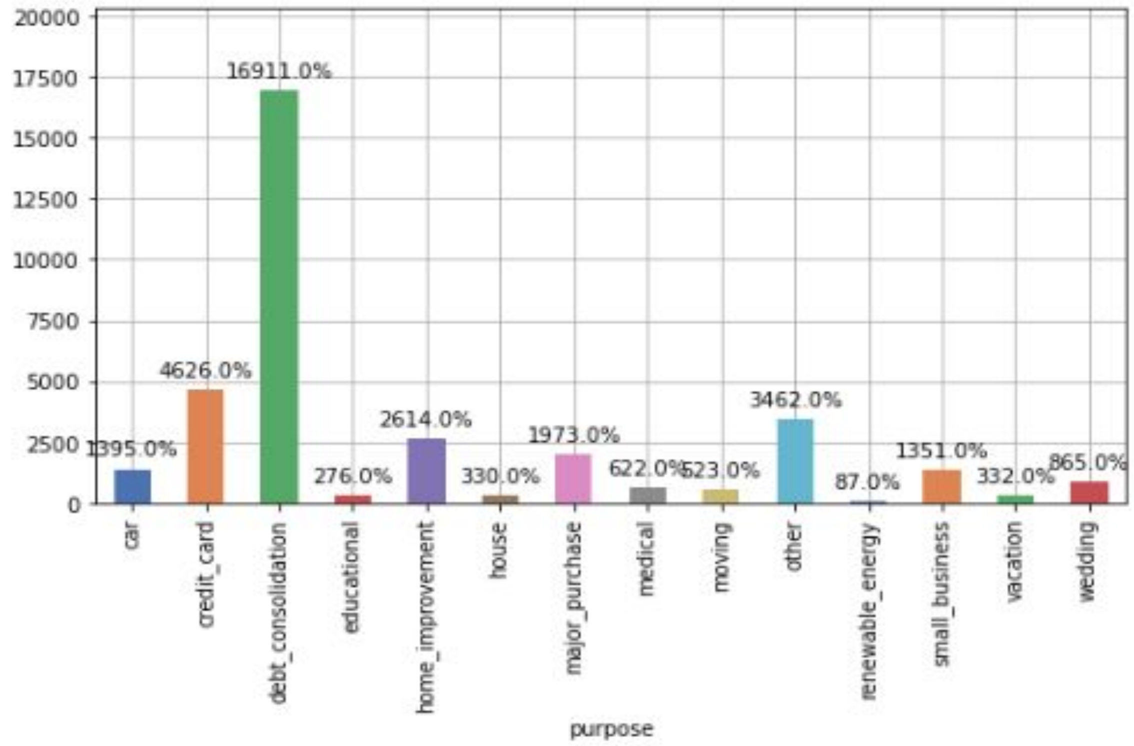
Lending Club only recovers 57% of the loan amount, when loans are defaulted.

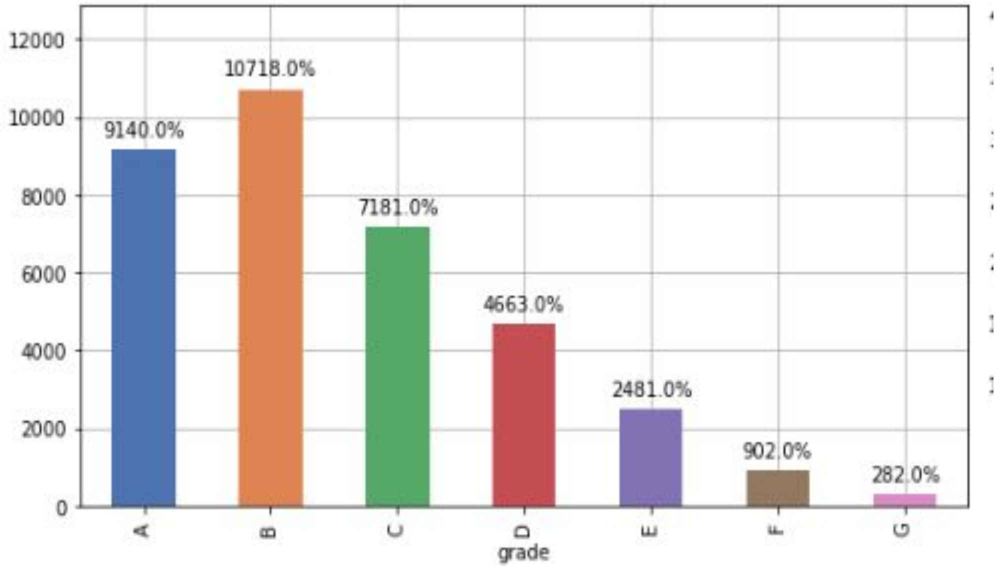
On fully paid up loans, the company makes 17% profit.



Analysis - Understanding Loan Details

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



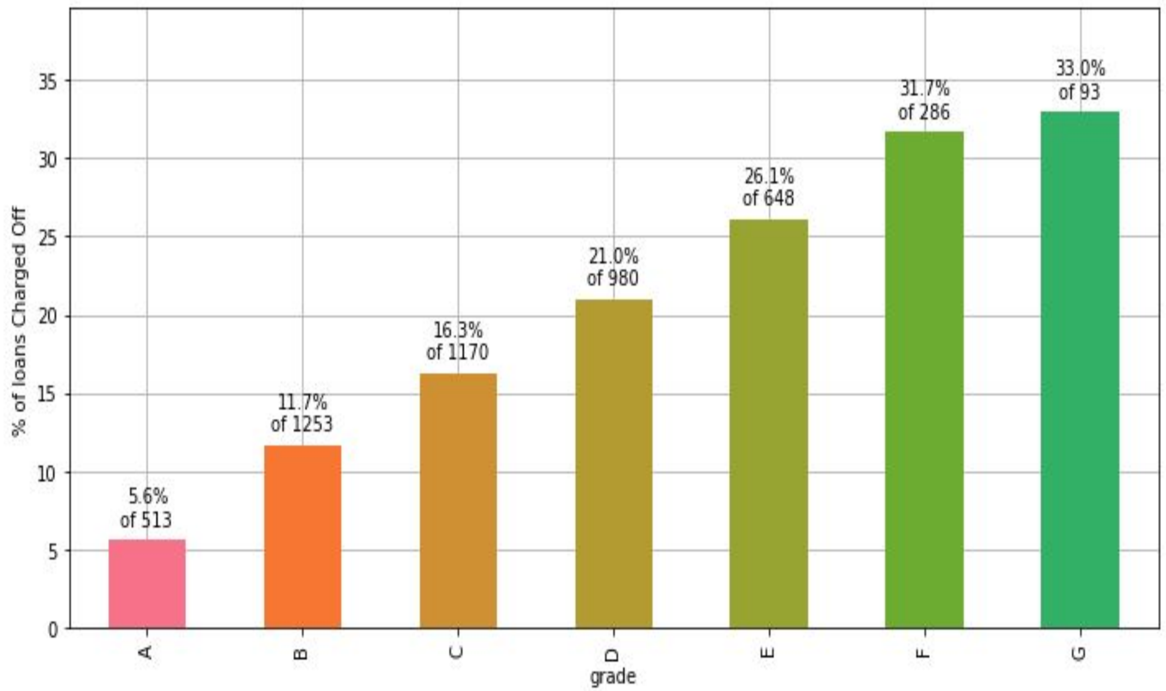


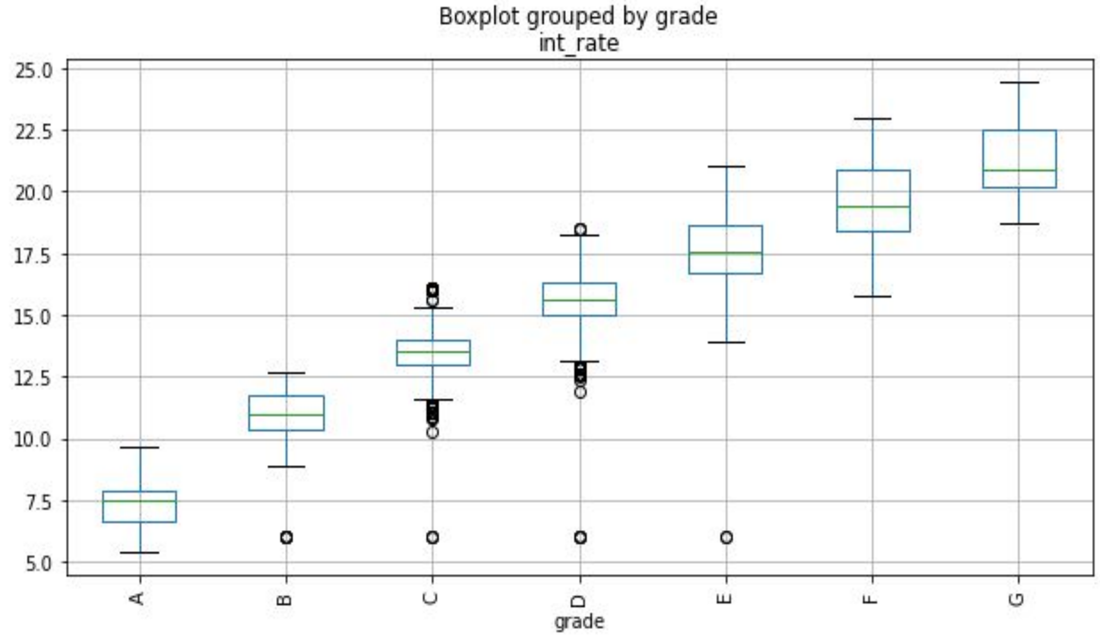
Most approved loans are high grade for A or B.

Analysis - Understanding Loan Details

Lower grades have higher incidence of defaults of loans. The grading system of Lending Club is working fine.

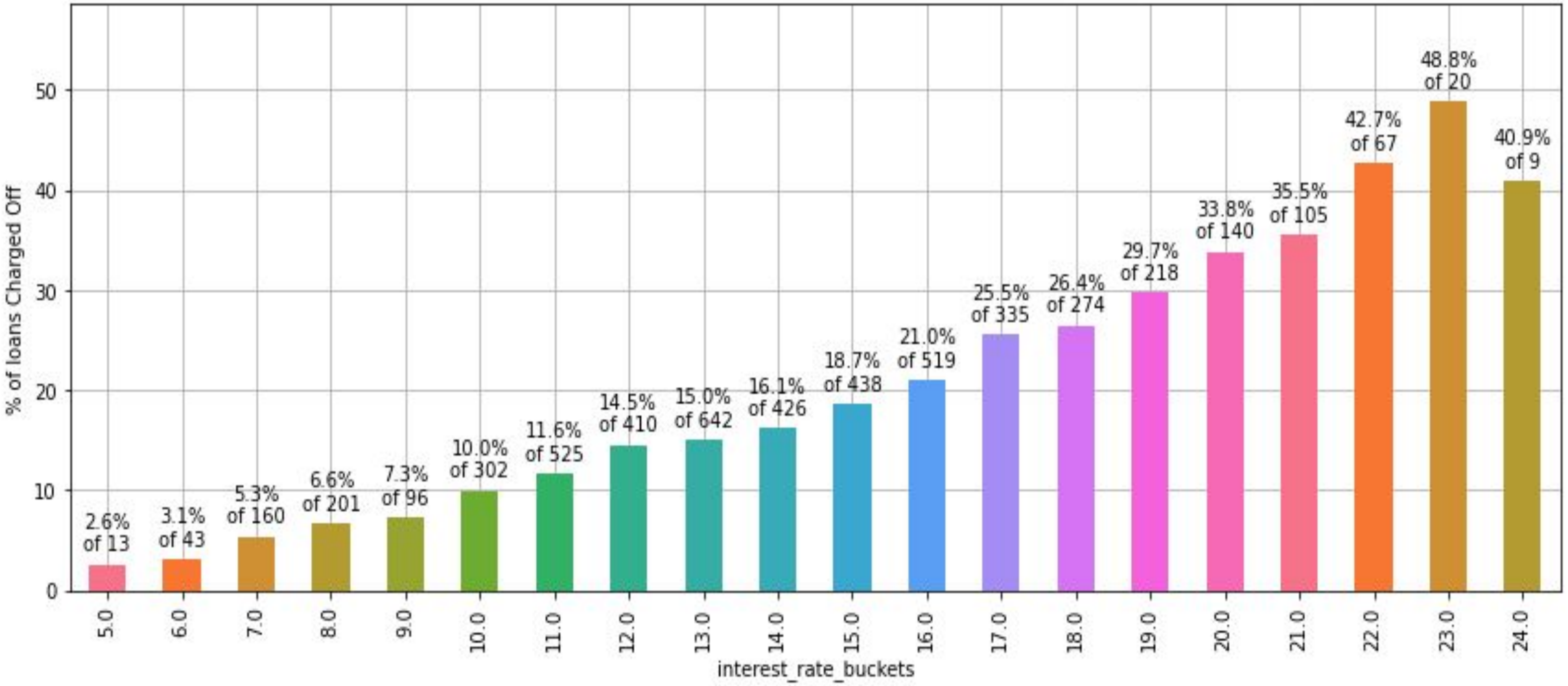
Lending Club charges higher interest rates as the grade of loans decrease. However we will see in next slides what are the driving variables for higher interest rates.

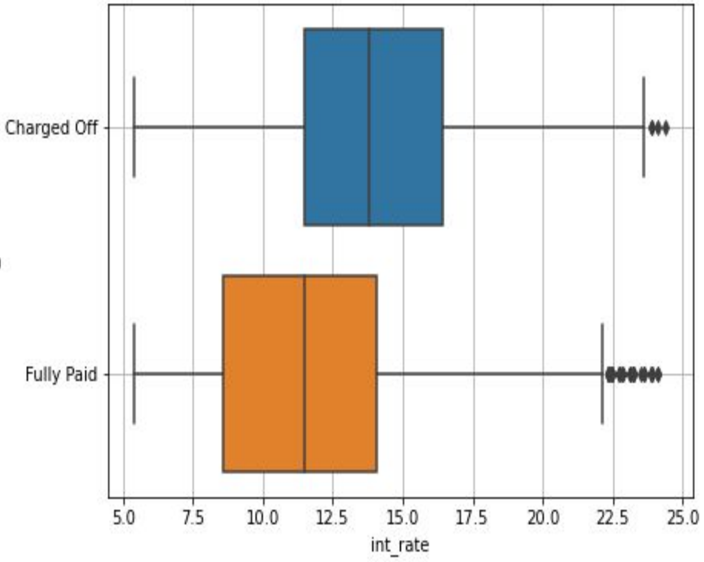




Analysis - Defaults by Interest Rates

Percentage of Defaults increases monotonically with higher interest rates. At a rate of 19% and above, more than 33% of loans are charged off.

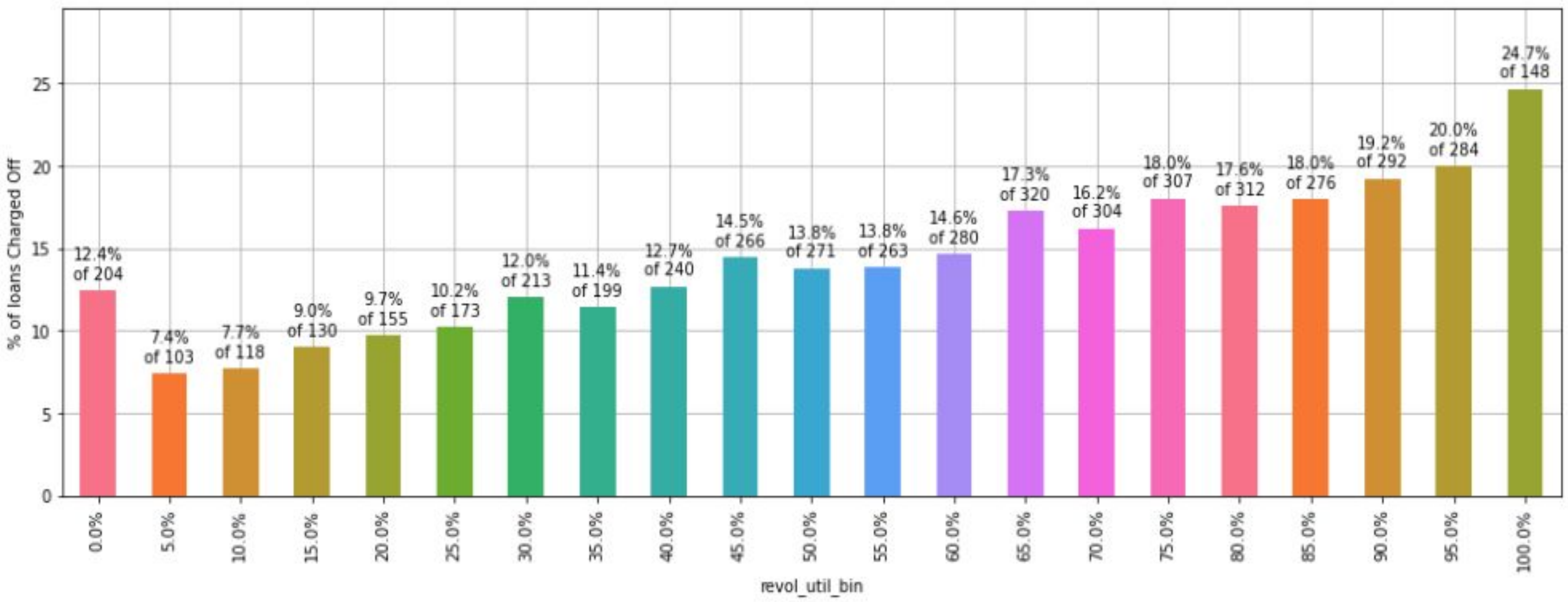


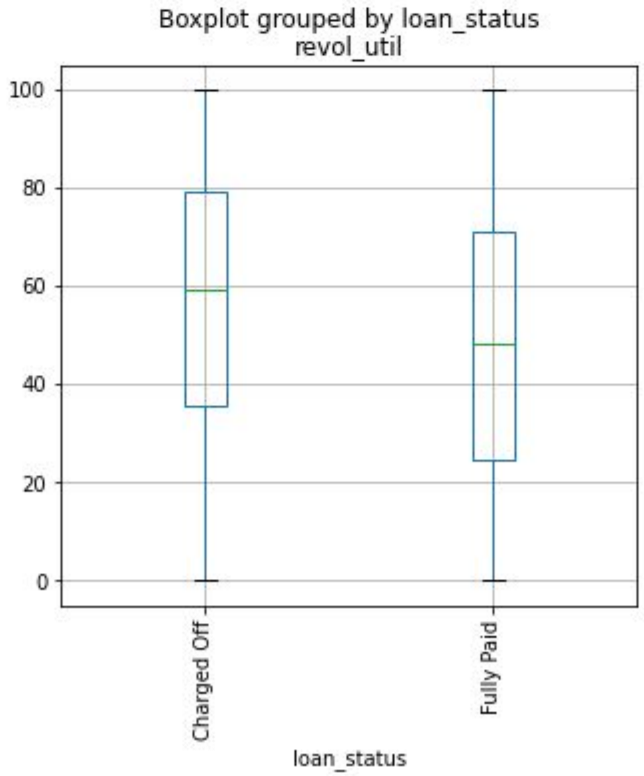


Analysis - Defaults by Revolving Line Util Rate

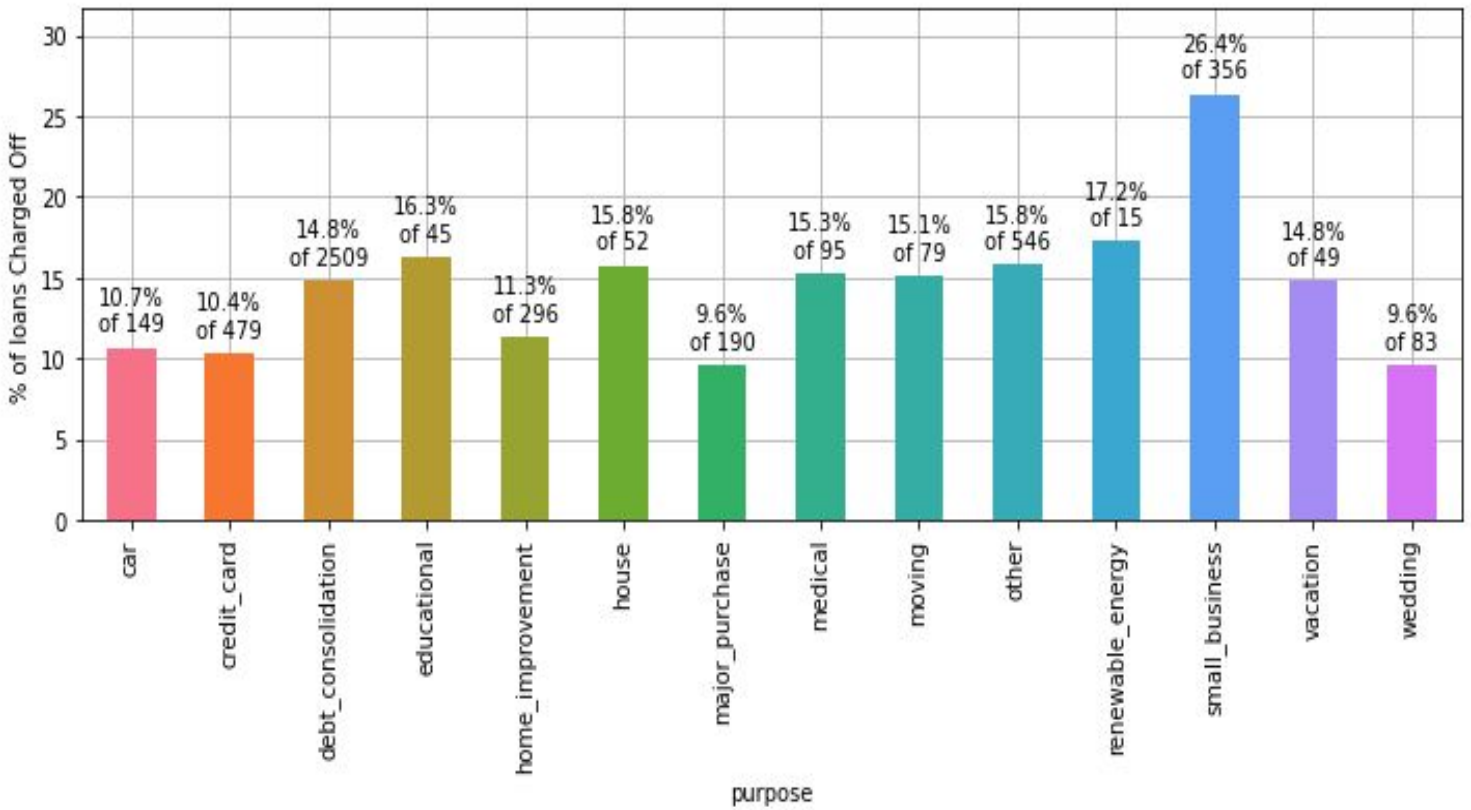
People with high utilization of Revolving Line of Credit at the time of taking loan default more.

Loans with utilization more than 75% are risky.





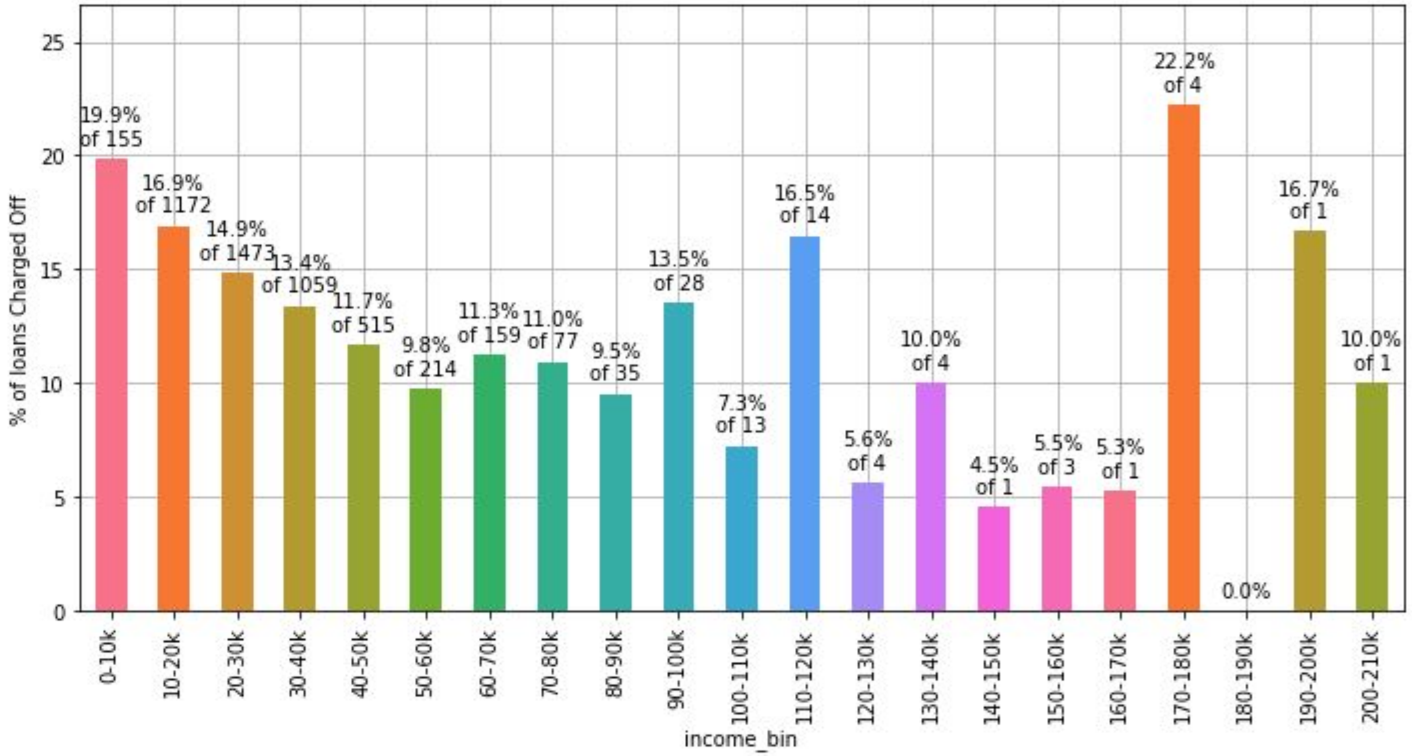
Analysis - Defaults by Loan Purpose



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

If you observe, the renewable_energy are charged off with 17% of the loans, which is too less.

Analysis - Defaults by Borrower's Income

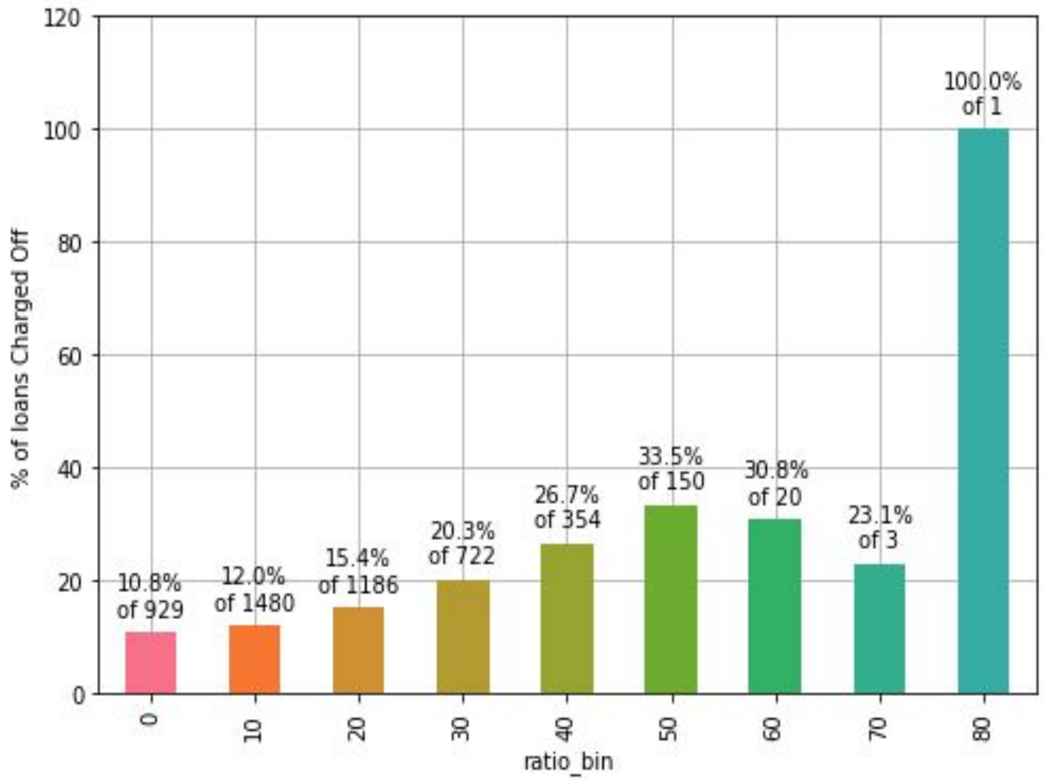


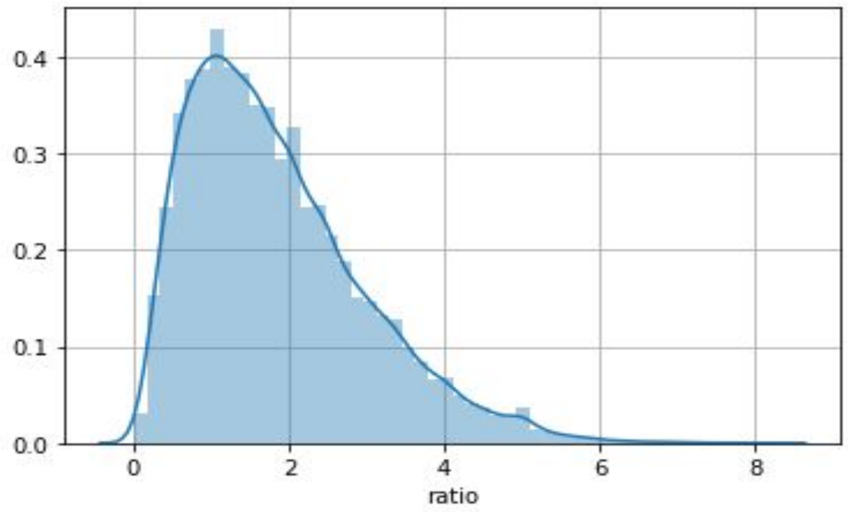
Borrower's having annual income of less than 20000 default on their loans at much higher rates.

Loan default decreases with higher annual income.

The ratio of amount to income is more important, which is presented in next slide.

Analysis - Defaults by ratio of amount to income

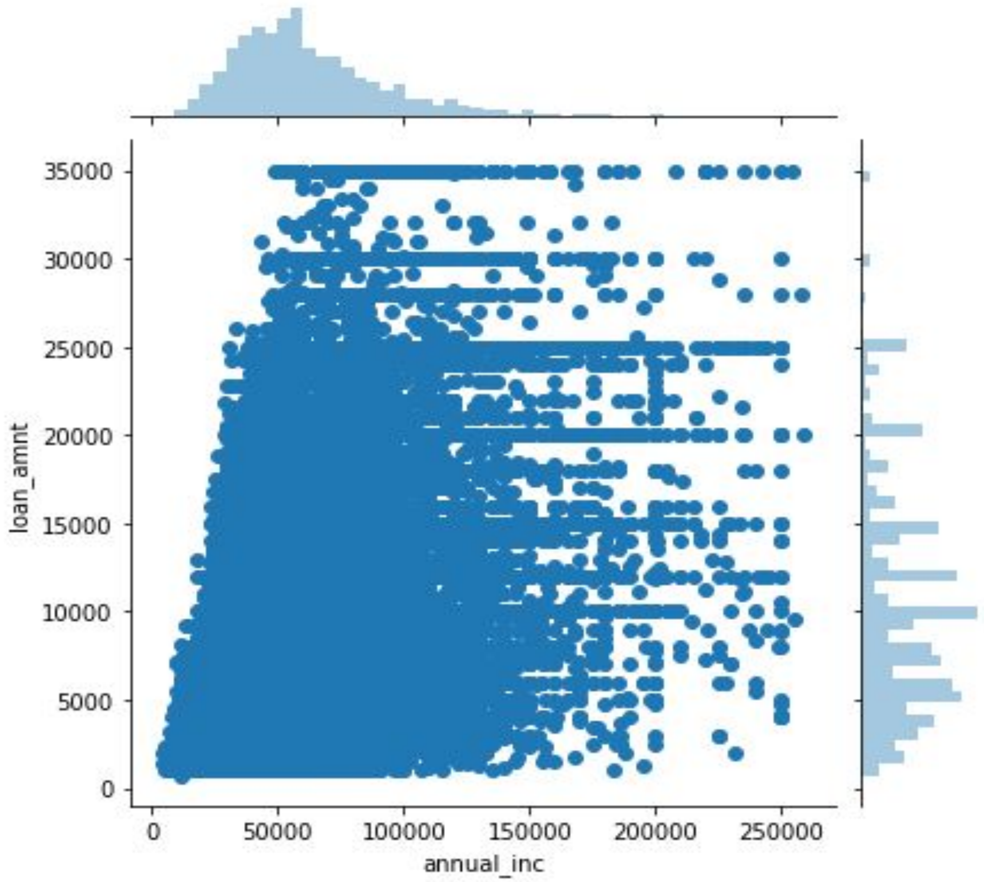




As long as loan amount is less than 20% of annual income, defaults are low.

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

Analysis - Defaults by ratio of amount to income



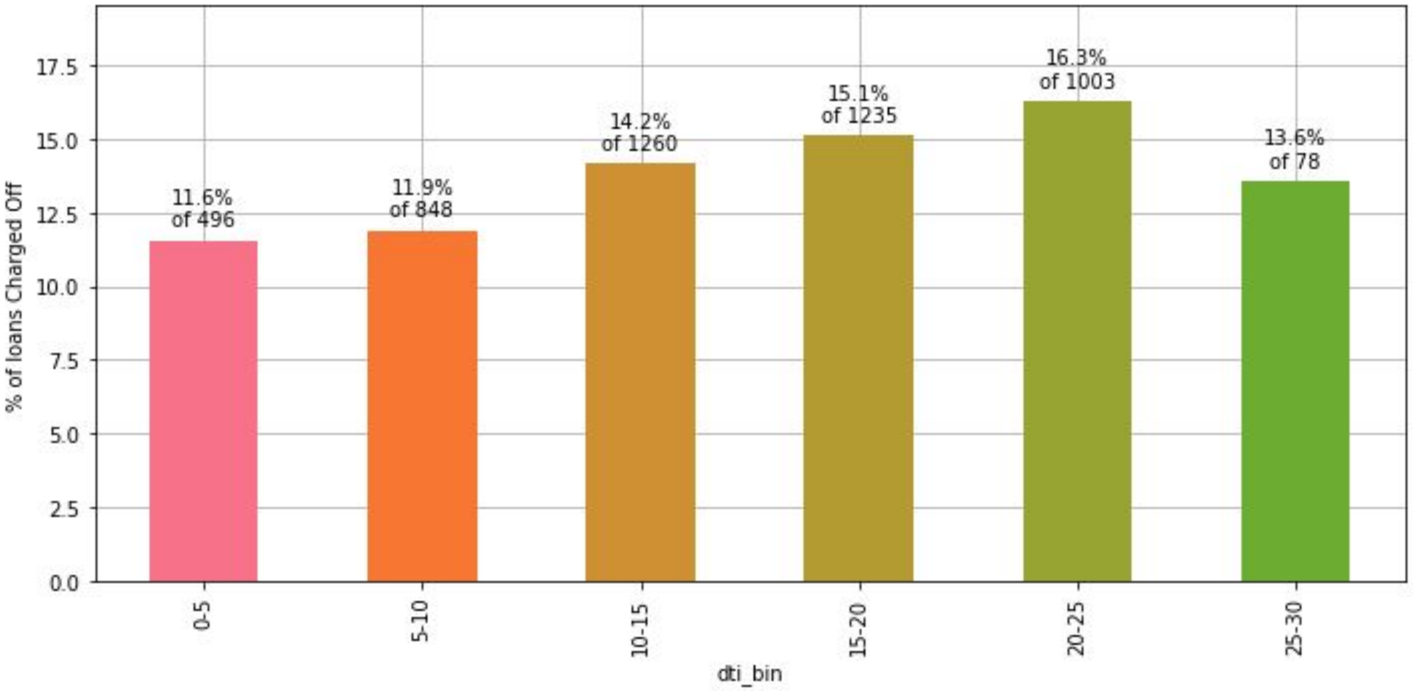
Lending Club has extended high-value loans to people with low income.

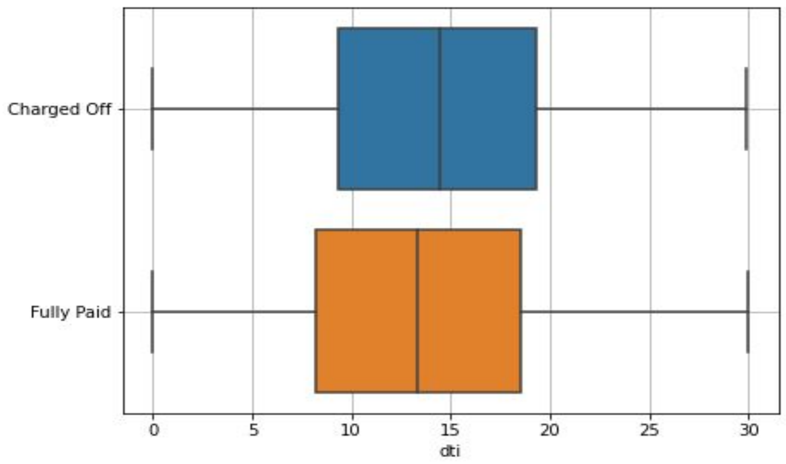
There are many cases of people with income with 50000 or less getting loans of 25000 or more. These would be risky loans.

This practice should be reduced.

Analysis - Defaults by Debt to Income ratio

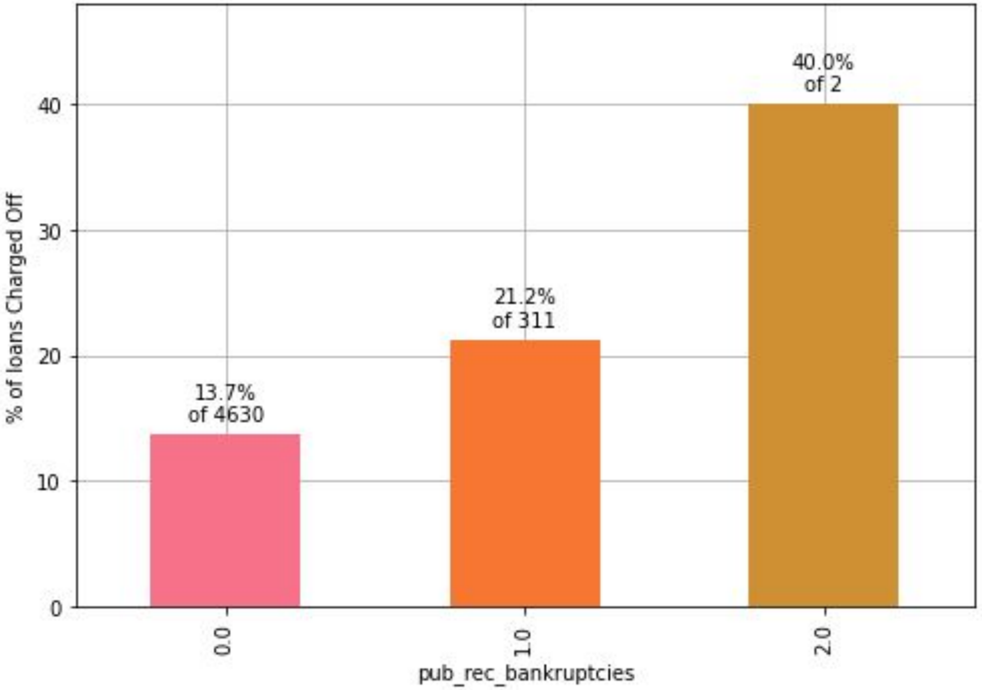
Percentage of default rises with dti ratio.
As the dti rises to 20% loans become risky.
Higher the dti higher the chances of loan being Charged Off



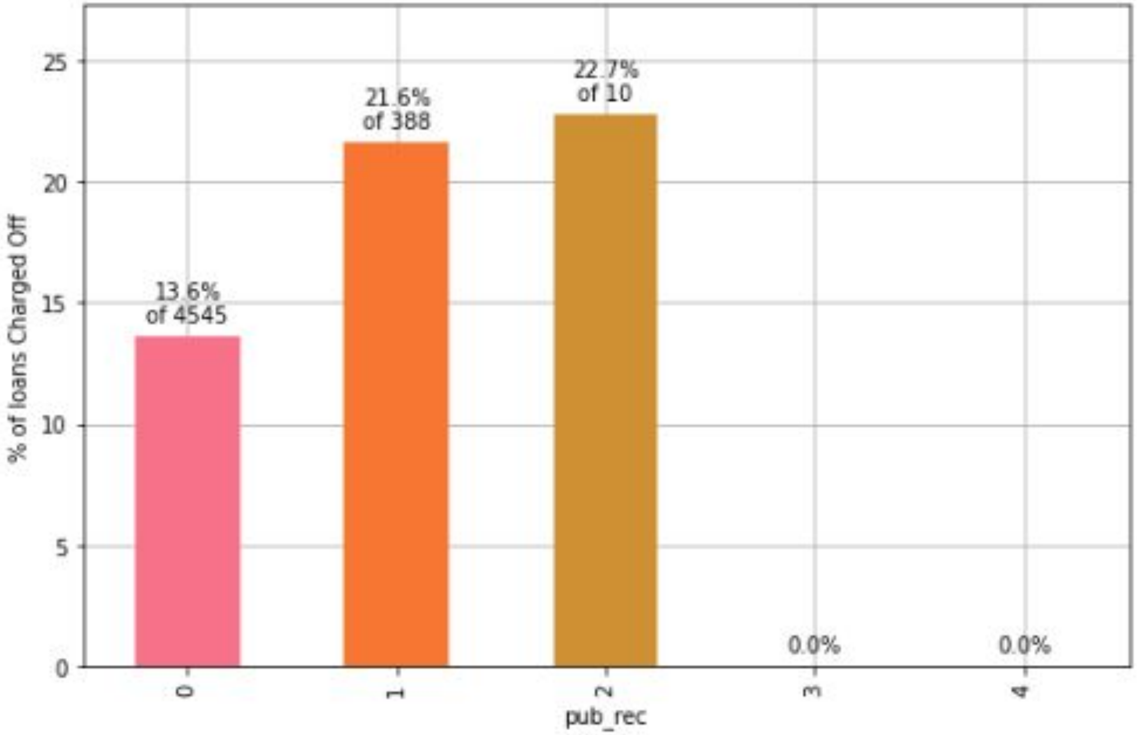


Analysis - Defaults by prior bad record

- 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. 5% have 1 derogatory record.
- Having even 1 derogatory record increases the chances of Charge Off significantly.

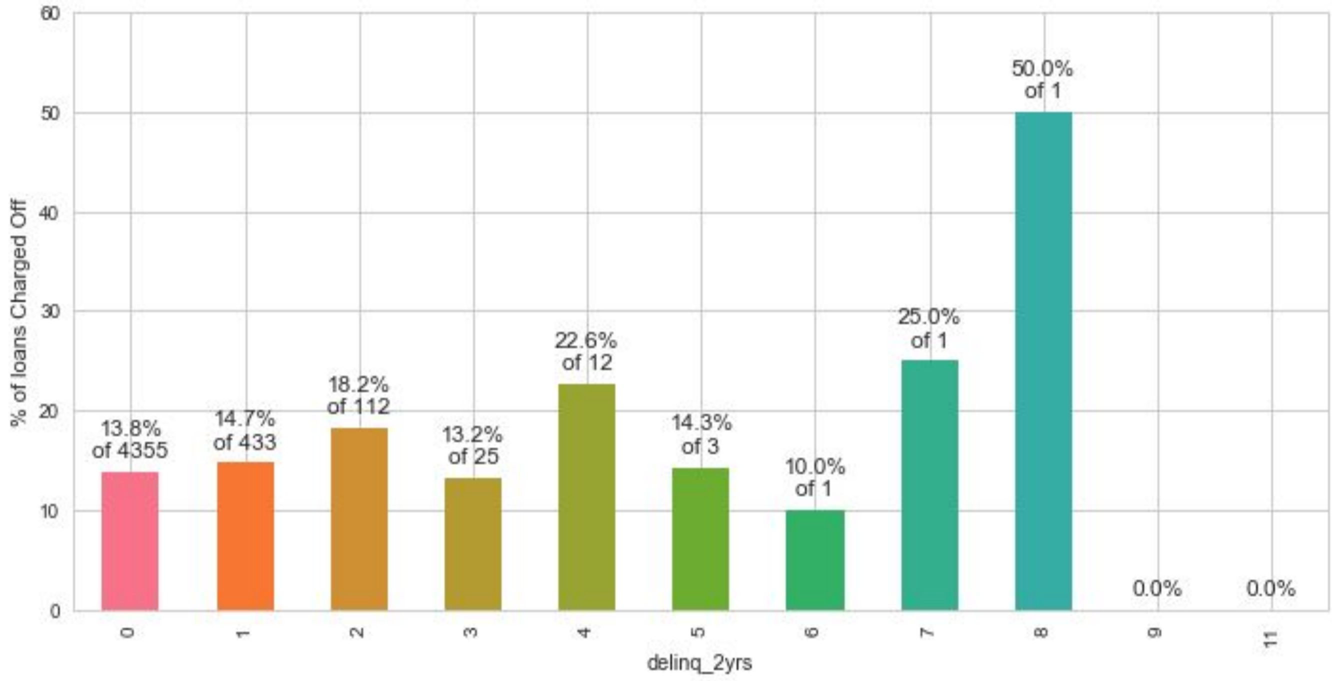


Recommendations

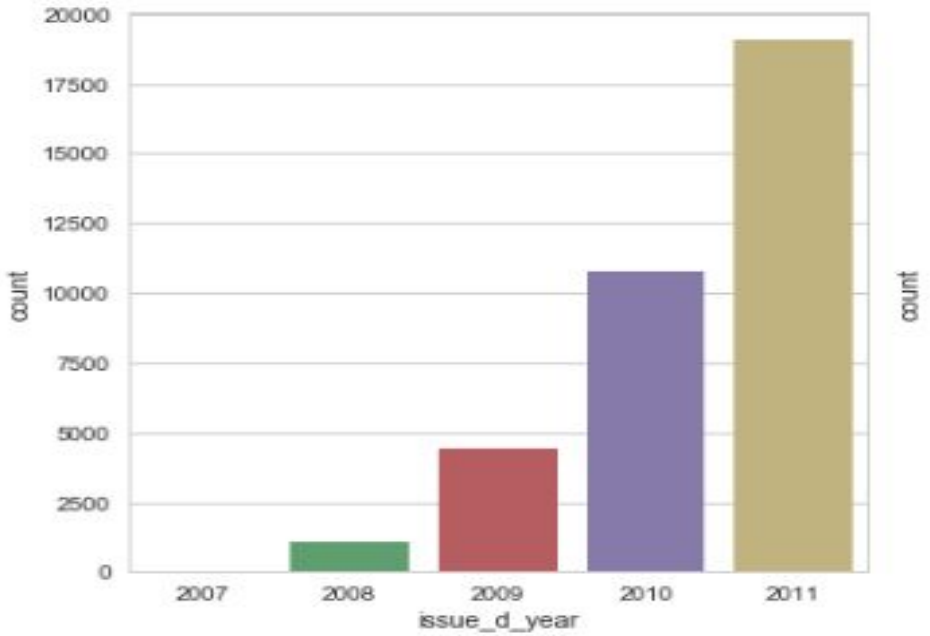
- **Stop approving loans where amount/income is higher than 30%**
- **Stop approving high-value loans when revolving line utilization rate greater than 75%**
- **Stop approving loans to people with prior bad record. Or at least stop approving high-value loans.**
- **Start charging higher interest rates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

Analysis - Interesting Insights

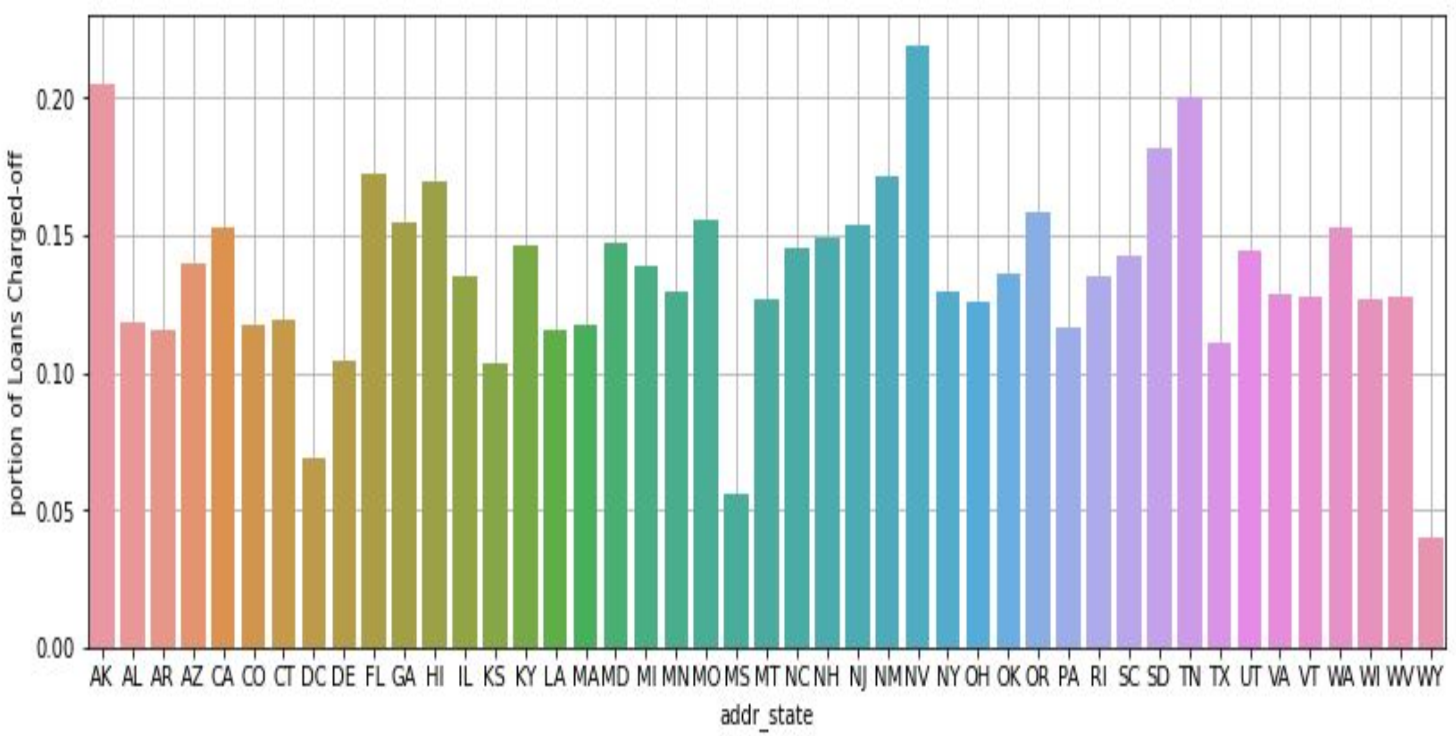
- The number of 30+ days past-due incidences of delinquency in the borrower's credit file for the past 2 years values ranges between 0 and 11.
- There are higher chances of default if this variable is greater than 1.



Over the years, Lending Club as an organisation has been giving more loans. From Aug,2007 to 2011 the loans issued have risen significantly.



Analysis - Interesting Insights

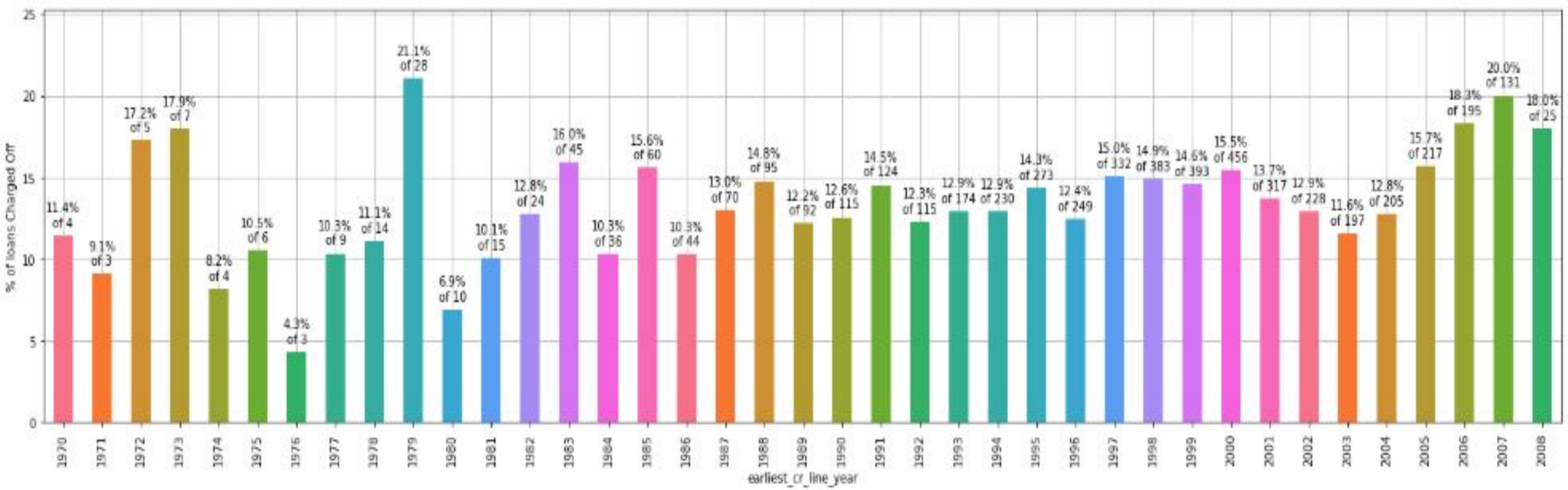


- Maximum loans ~18% are from California following that is 9.5% from New york state and Florida with 7%. This is to be expected as these are also the three most populous US states
- States with higher Charge Off rates have very low number of loans. The percentage is therefore NOT significant and should be ignored. Overall, this variable does not affect propensity to default.

Analysis - Interesting Insights

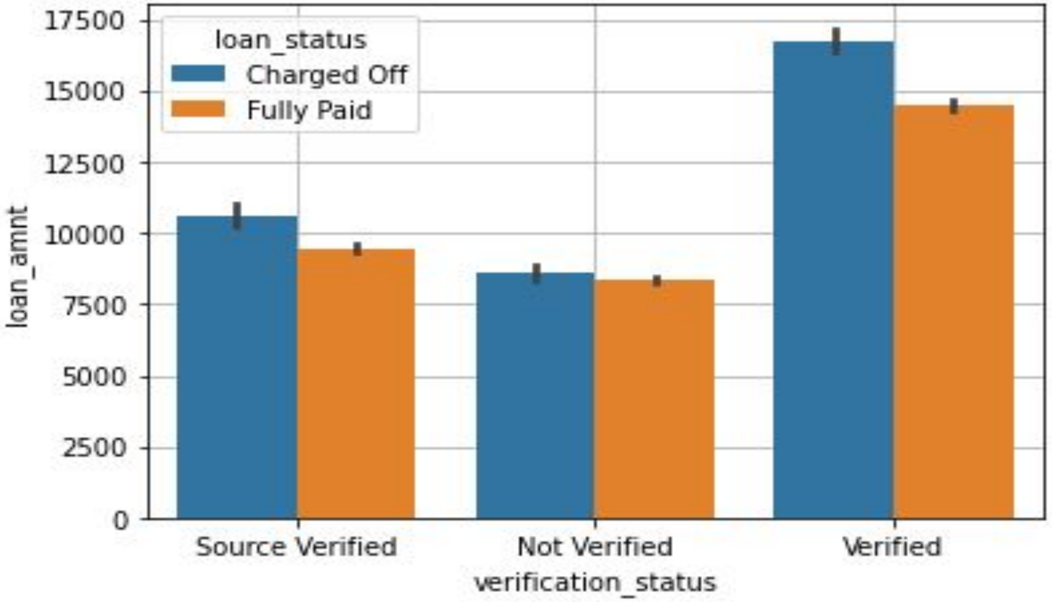
People who took their first loans just before an economic crisis, like the one in 1980 and the subprime crisis in 2008, have higher rates of default.

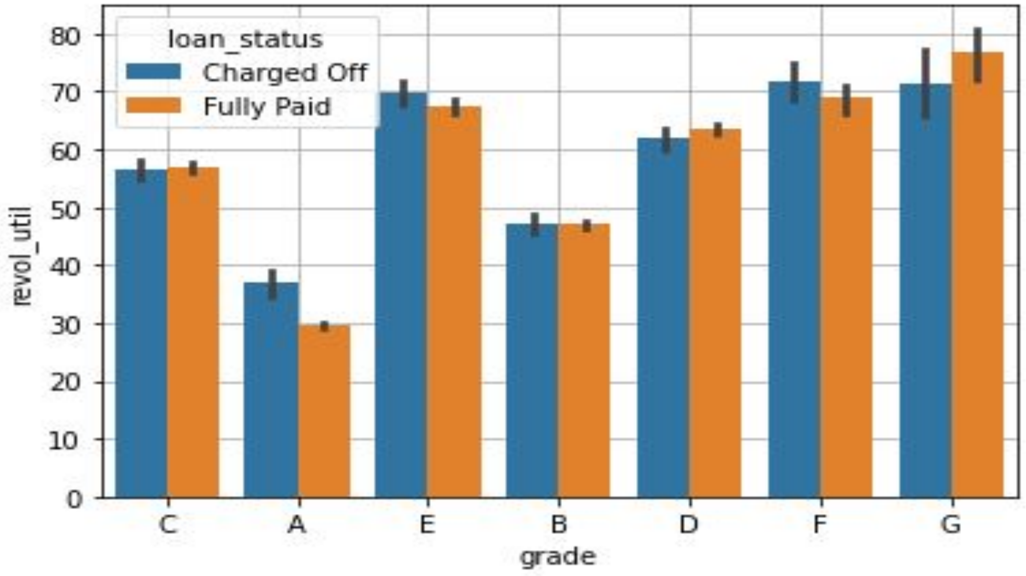
Presumably, these are young people who were affected by the economic conditions of their early career



Analysis - Interesting Insights

- Higher loan amounts are Verified more often.
- This is why verified loans see a higher rate of default. It's not the verified status., it's the fact that higher loan amounts are riskier and are also verified more often by Lending Club.





- revol_util and grade(and therefore int_rate) are correlated in some way. The revol_util is positively correlated to the grade.
- As the grade goes from A to E the revol_util also increases. This may be because higher loan amounts are associated with higher grades.

Analysis - Interesting Insights

Not many loans receive investment with higher number of delinquencies(>4). Despite the low loan amount request, these loans are considered risky and are not invested much in.

Lending club should further restrict their investment. We see loan amounts of >15000 on average for people having>=2 delinquencies.