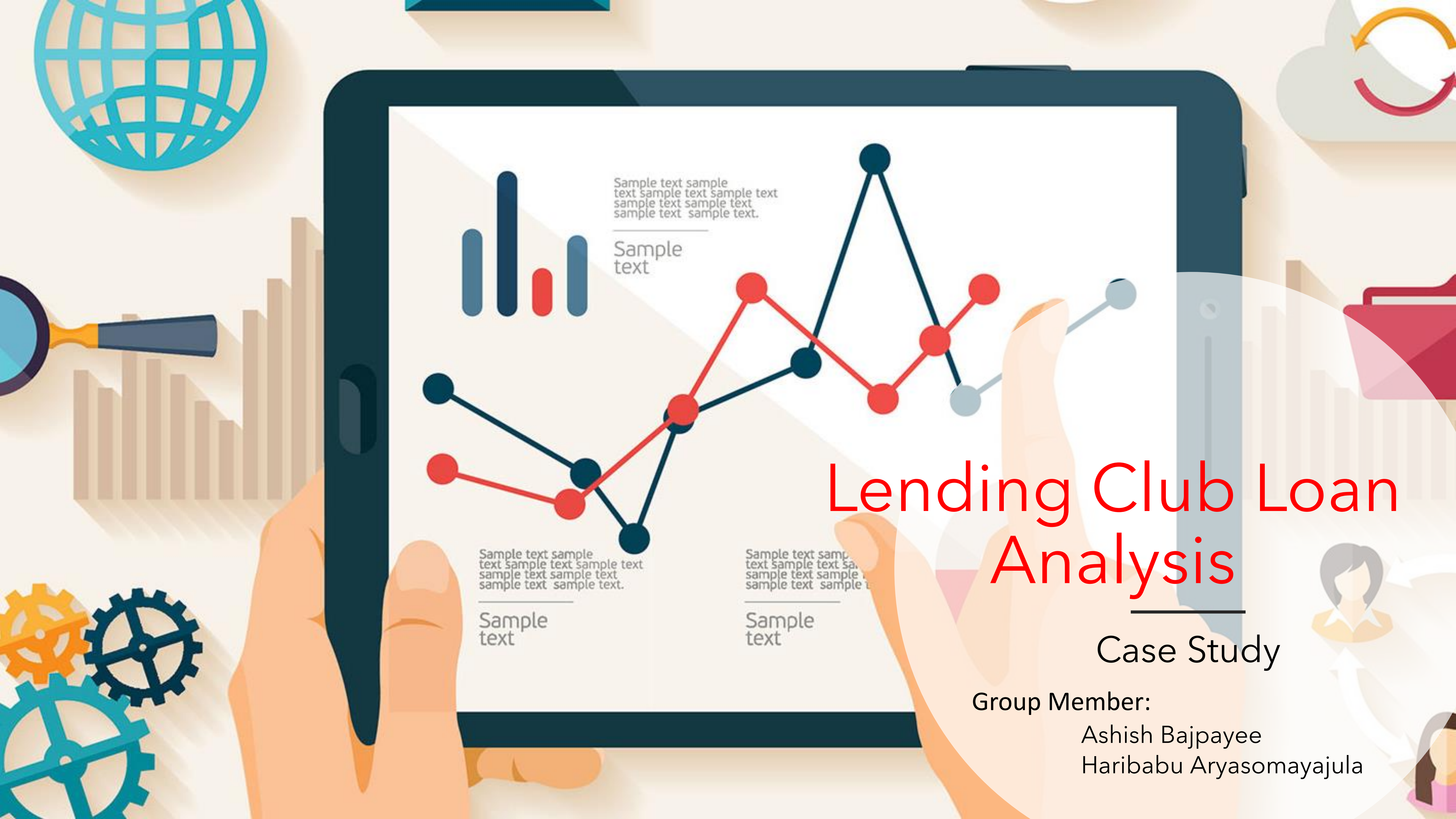


Lending Club Loan Analysis

Case Study

Group Member:
Ashish Bajpayee
Haribabu Aryasomayajula

Overview

upGrad

Company History

Since 2007, Lending club has been bringing borrowers and investors together, transforming the way people access credit. Over the last 10 years, They have helped millions of people take control of their debt, grow their small businesses, and invest for the future..

Core Business

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 quick online interface.

Business Objective

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.

Problem Statement

By applying EDA technique, develop understanding of risk analytics in banking and financial services and understand how data is used to minimize the risk of losing money while lending to customers.

Analysis Approach

upGrad

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

Data Wrangling

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 Univariate Analysis

Bivariate Analysis

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

- ✓ *Publish Insight analysis report*

Conclusion

Input and Assumption

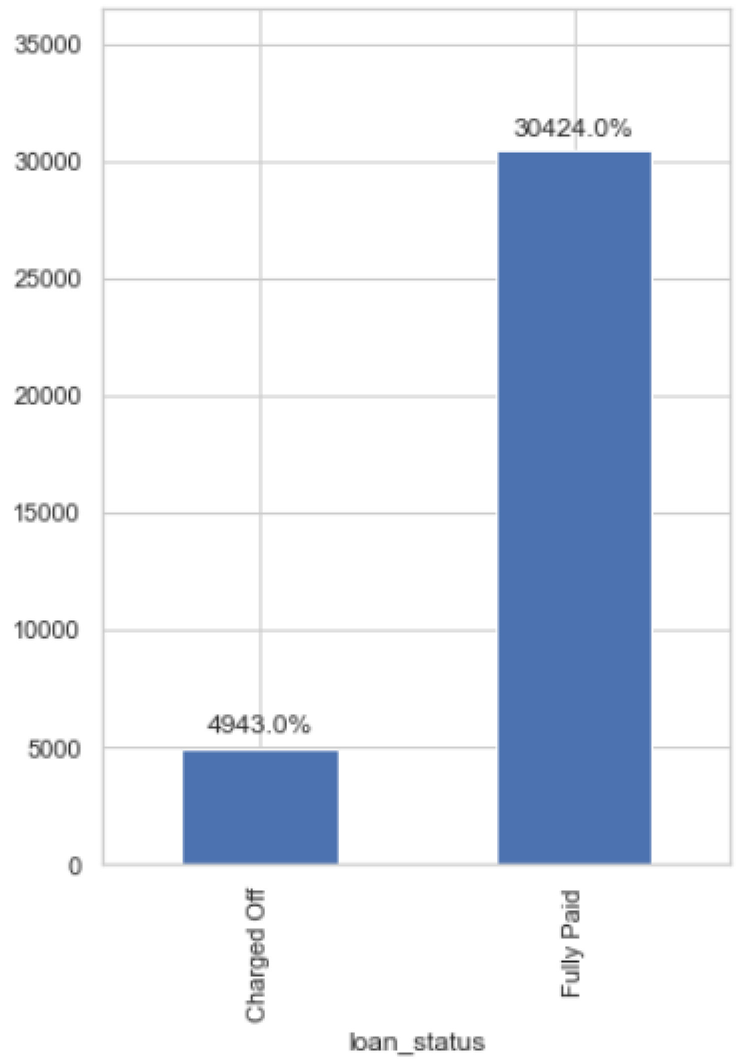
Source Data:
Following source data has been provided to us for analysis:

- Private Loan Data of Lending Club through the time period 2007 to 2011.
- Data dictionary of defined variables.

Assumption:
- Decision on loan granting depend up on loan request if graded by LC and then listed for loan.
- It is up to the investor who decides which loan listing to invest in looking at the loan detail and borrower's information.
- There are only those variables be chosen for analysis that will be available to the investor at the time of deciding whether to invest in a loan request or no

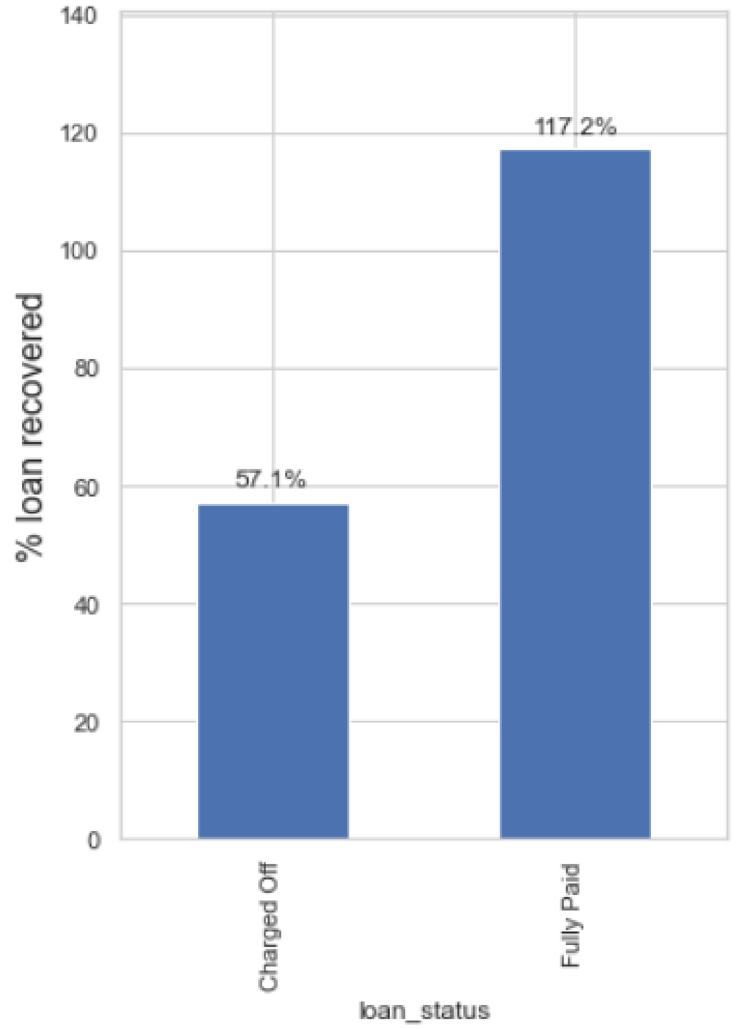
First glance Analysis

Total Loans



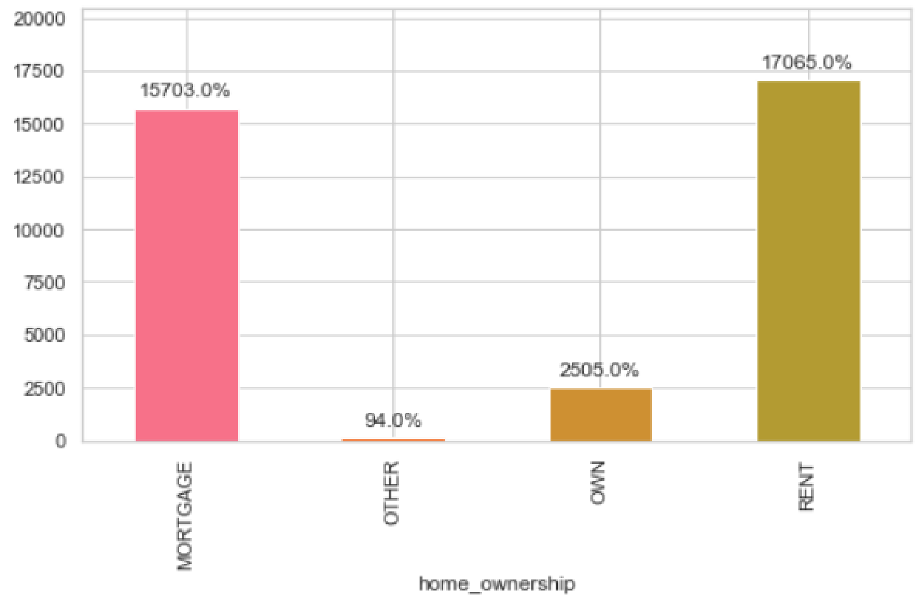
- Approximately **14%** of loans are defaulted
- Any variable that increases percentage of default to higher than 16.5% should be considered a business risk. (16.5 is 18% higher than 13.97 - a large enough increase)

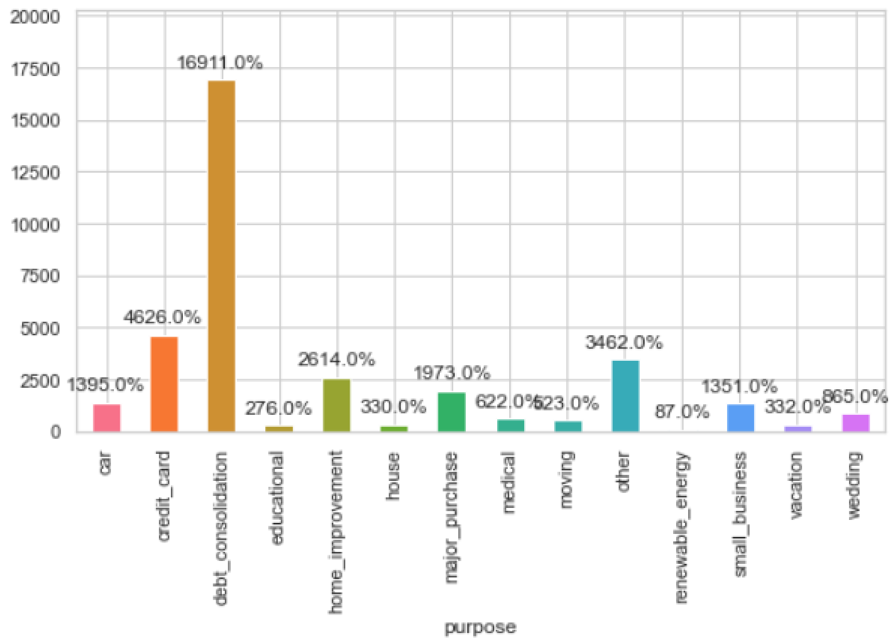
Total Money Earned

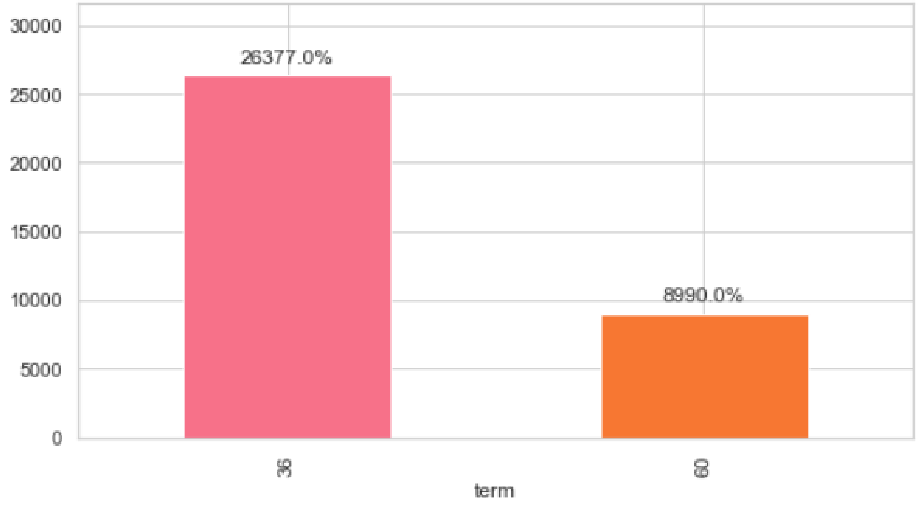


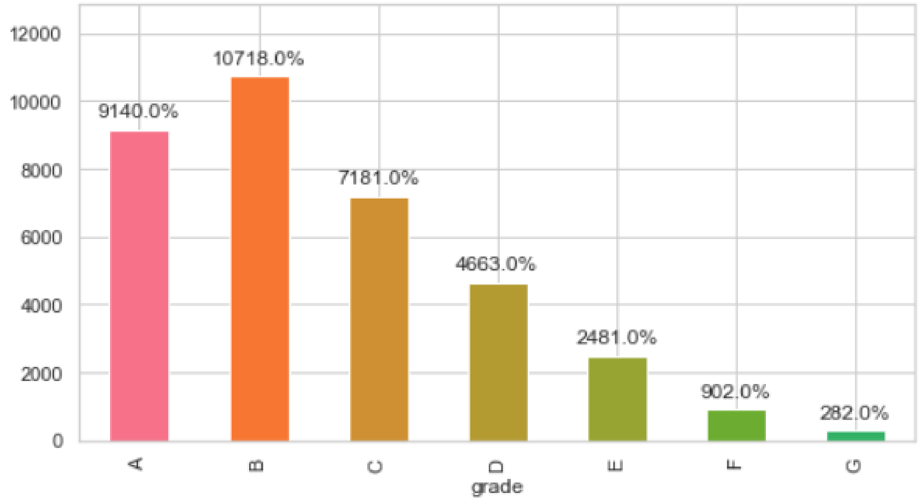
- Lending Club only recovers **57%** of the loan amount when loans are defaulted.
- On fully paid up loans, the company makes **17%** profit.

Loans Understanding - Analysis





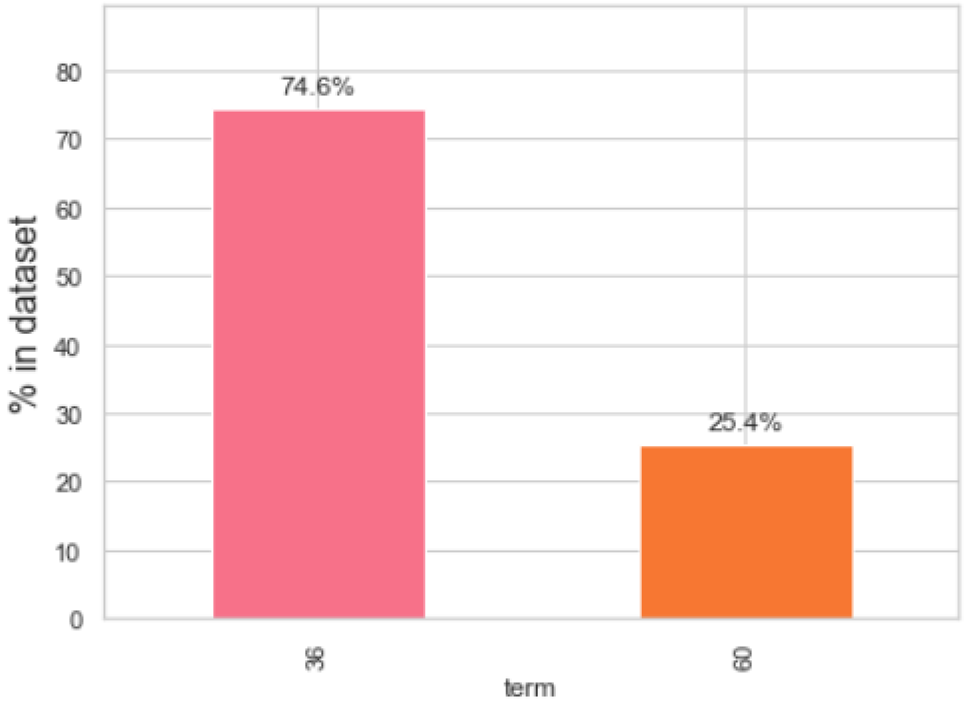


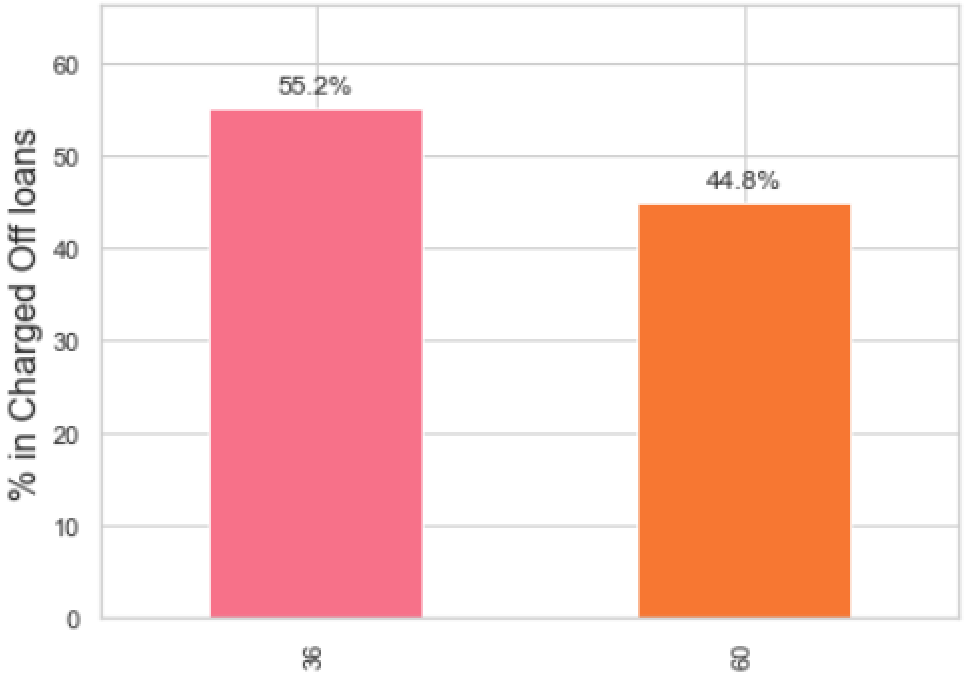


Observations:

- Maximum number of borrower home status ownership are **RENT** followed by Mortgage.
- **Debt Consolidation** is the largest 'purpose' for seeking a loan followed by **credit card**
- Most loans are for **36-month** term
- Most approved loans are **high grade**
- Focus of investigate the frequencies of defaults by **term, purpose and grade**

Terms- Analysis

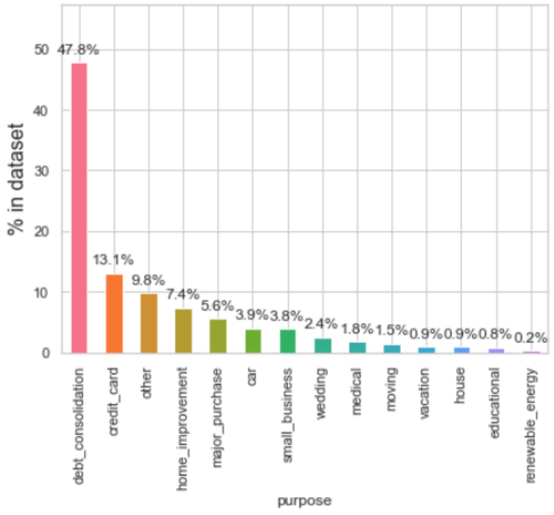


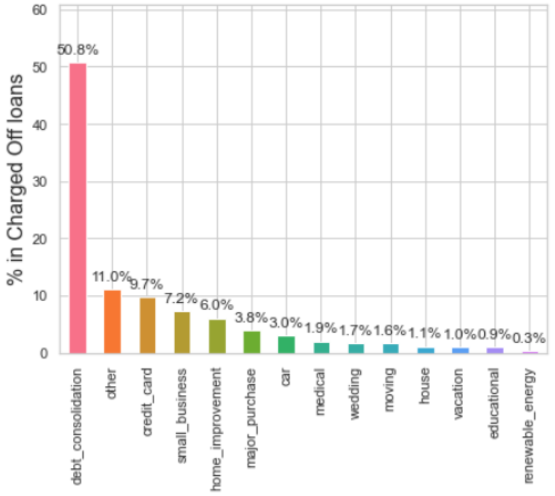


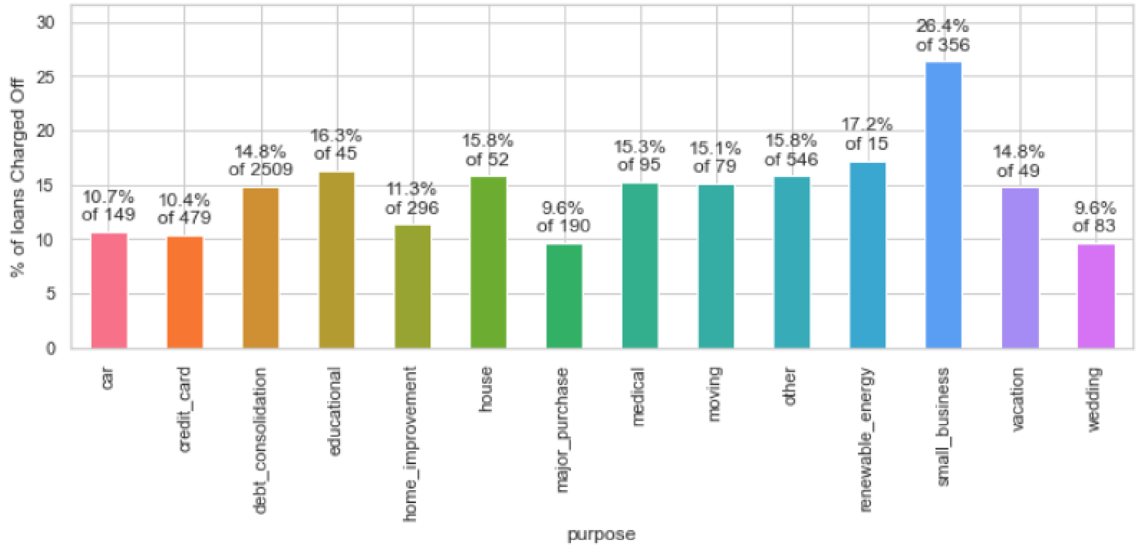
Observation:

- ✓ Around 75% of the total loans are given for duration of 3 years. while just 25% of the loans are those given for 5 years.
- ✓ Among Charged Off loans, percentage of term 60 months rises to 45%. The higher term loans have a higher chance of default.

Purpose- Analysis



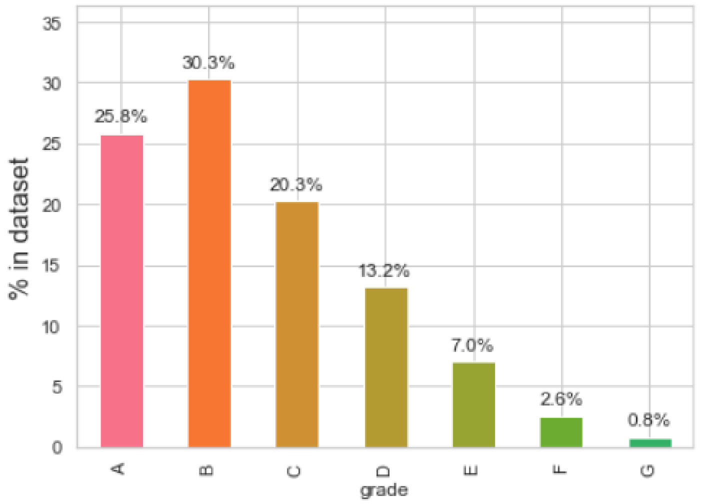


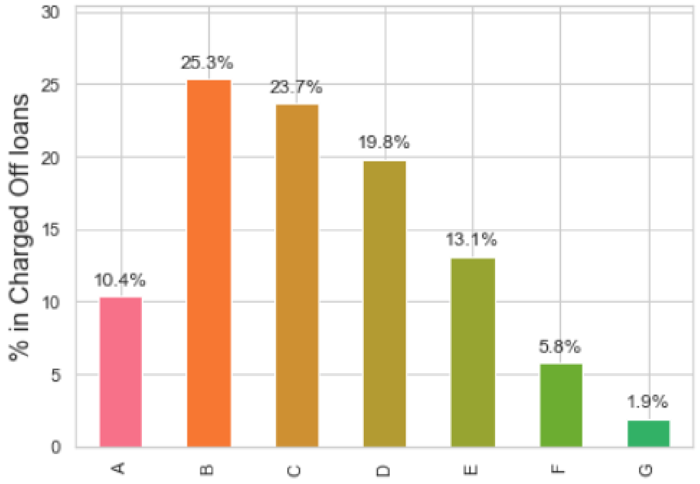


Observation:

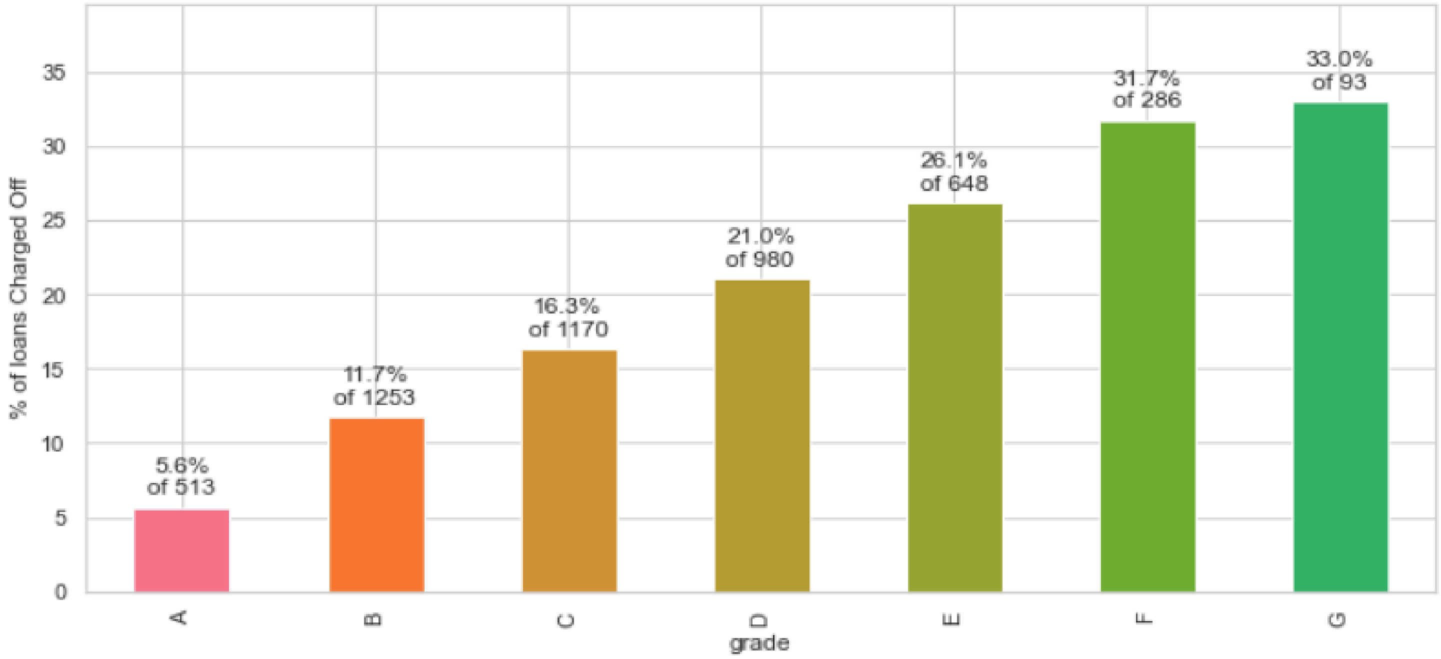
- ✓ The category 'small business' percentage doubles from 3.8 to 7.2 for Charged Off loans.
- ✓ **26%** of loans for small business are Charged Off. Making them the riskiest purpose.
- ✓ Approximately **~49%** of the loans are issued for the purpose of dept consolidation.
- ✓ 17% of the loans for renewable energy are charged Off, but the number is too less to be of significance.

Grade - Analysis





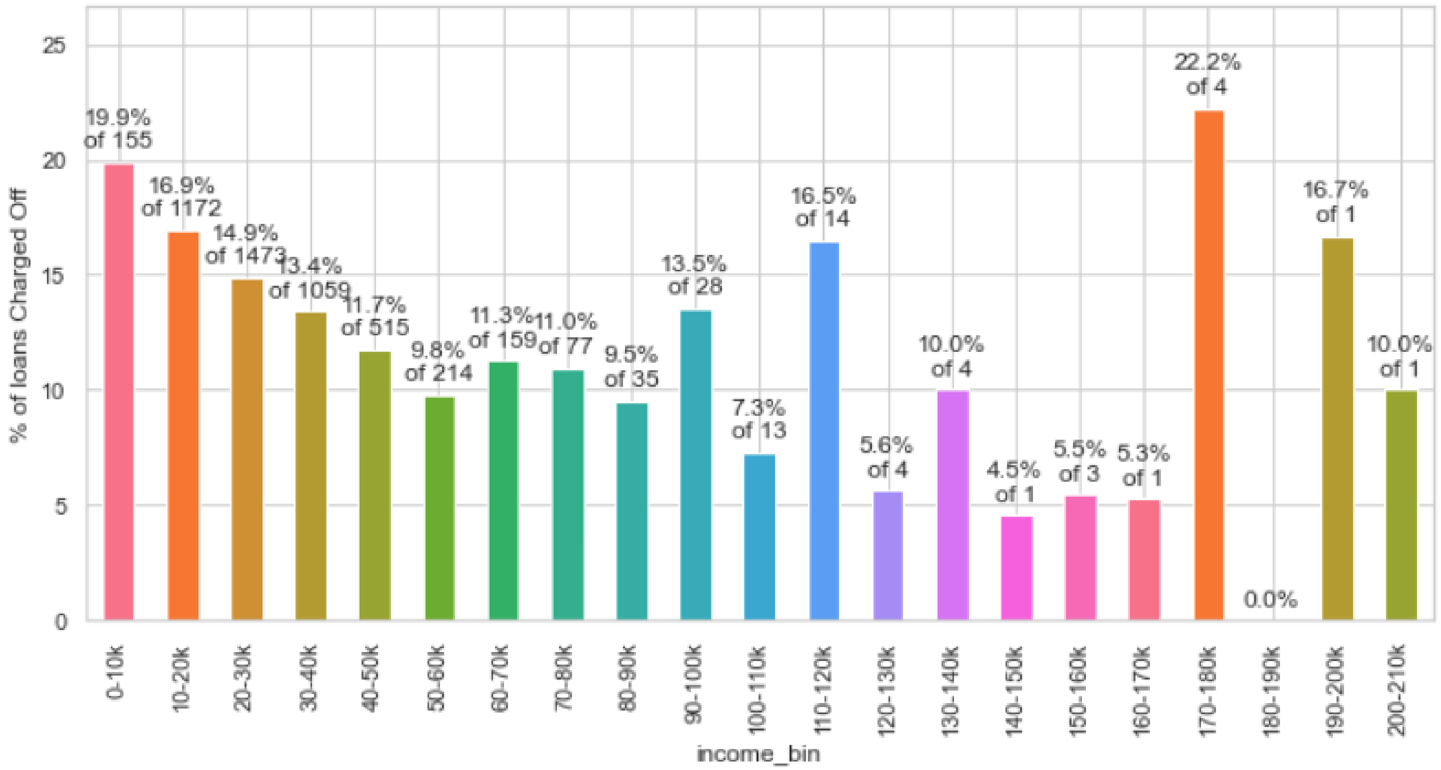
Grade Contribution analysis

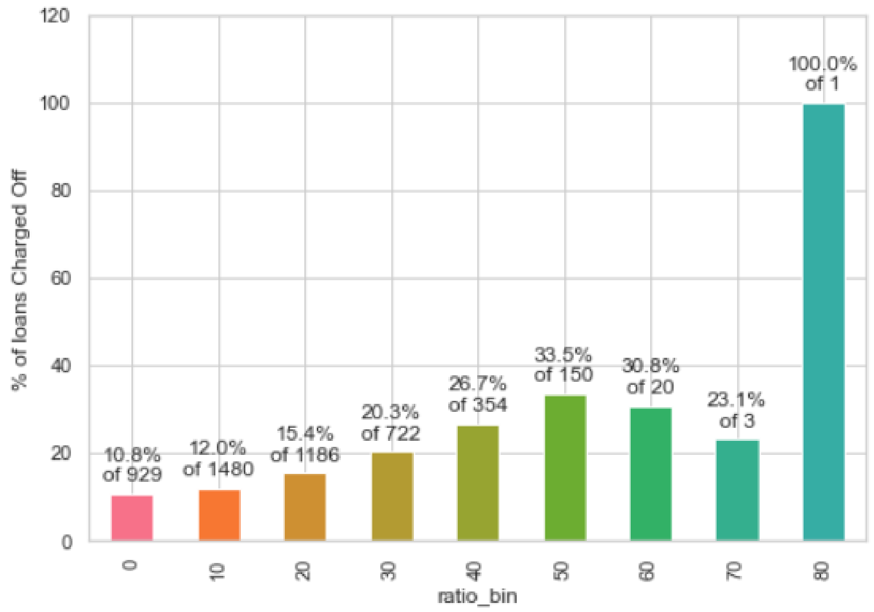


Observation:

- ✓ **Grade A and B loans are safe**. The percentages in full dataset are much higher than percentages in Charged Off loans.
- ✓ **Grade D, E, F, G loans are less safe**. We should plot grade by percentage Charged Off by category
- **Lending Clubs grading system is working well.**
- Nearly 30% of all loans in Grades F and G see a default.
- Grade E onwards are risky, and less numerous. Lending Club should either refuse loans or charge high interest rates.

Annual Income - Analysis

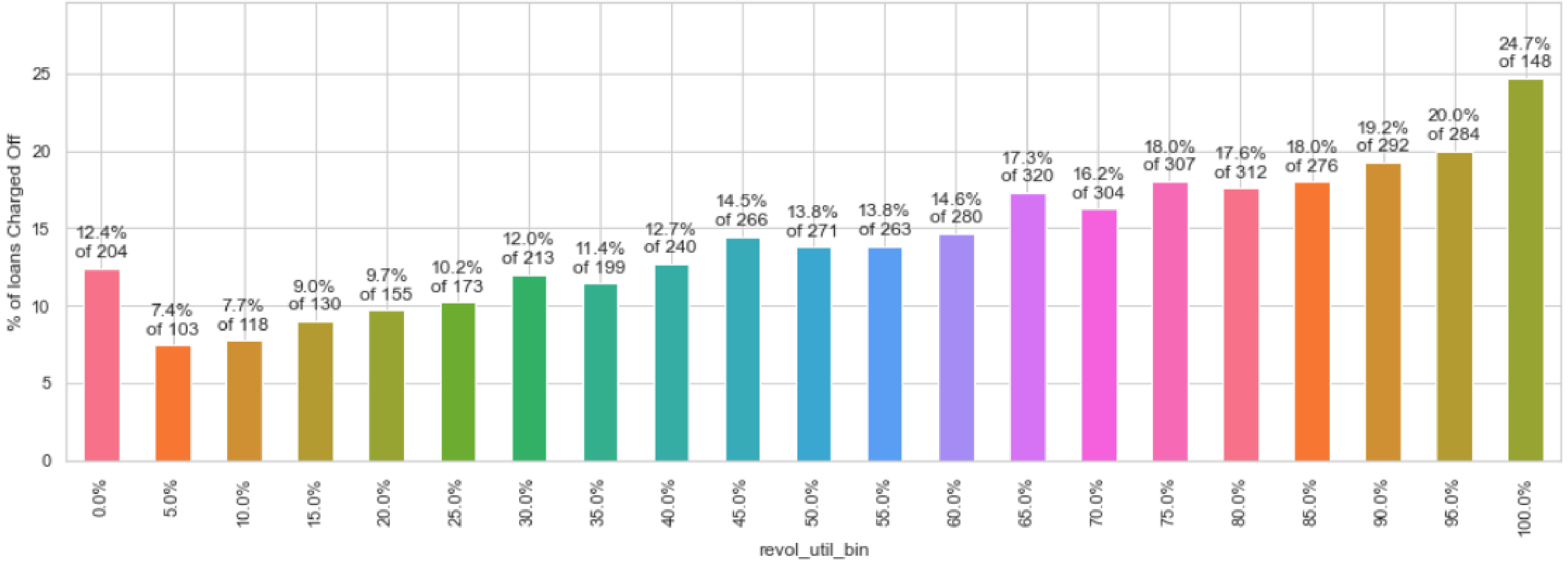


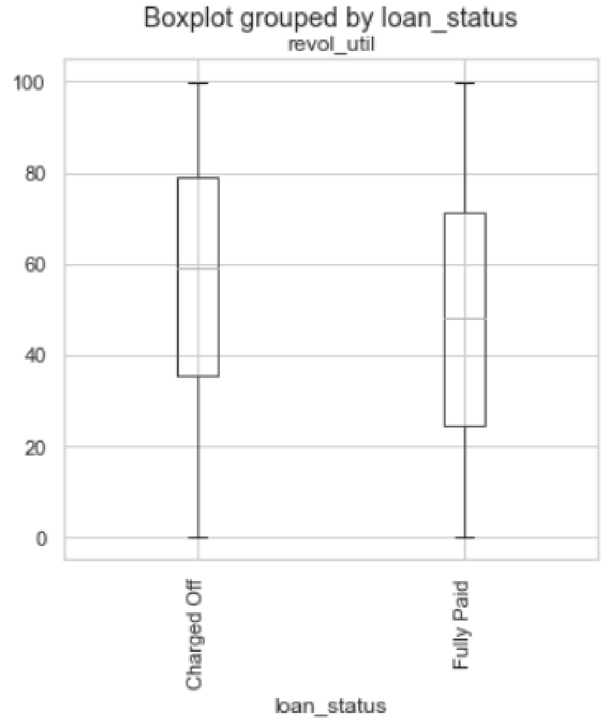


Observation:

- ✓ Loan defaults are higher for lower income, and progressively reduce as incomes go up.
- ✓ If 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.

Revolving utilization rate- Analysis

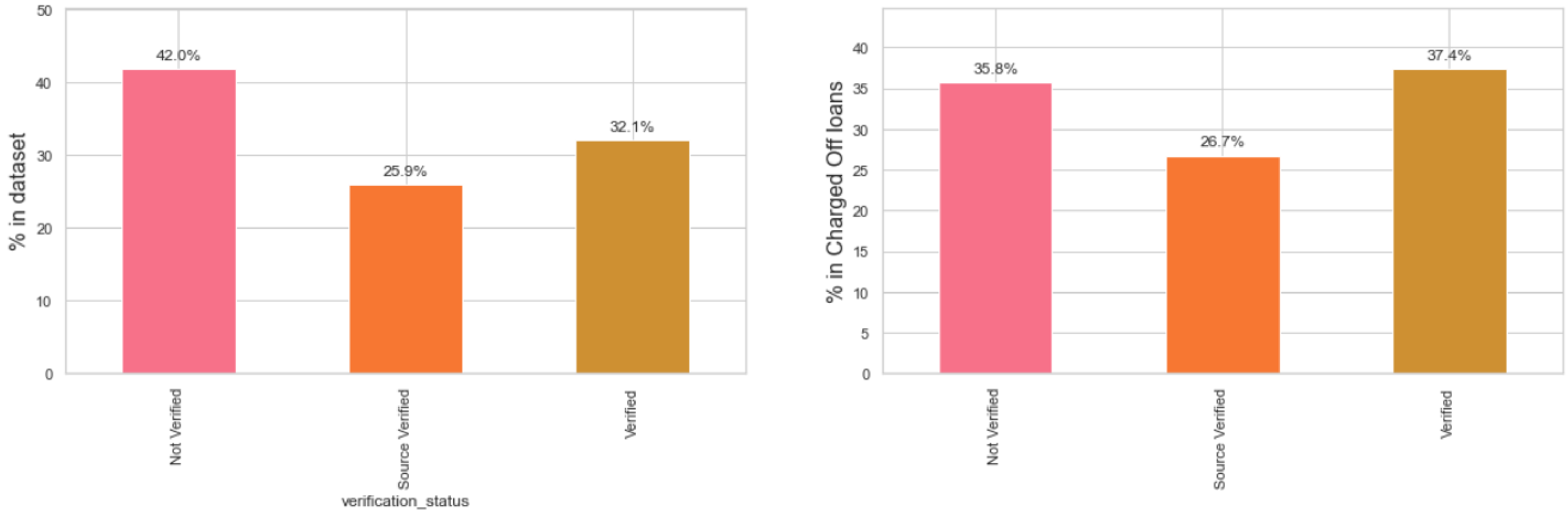




Observation:

- ✓ Average revolving utilization rate is seen to be higher for Charged Off loans than the Fully Paid loans.
- ✓ People with high utilization of Revolving Line of Credit at the time of taking loan default more.
- ✓ Loans with utilization > **75%** are risky. This practice should be stopped.
- ✓ Density of low value loans is also high. They should be approved less often.

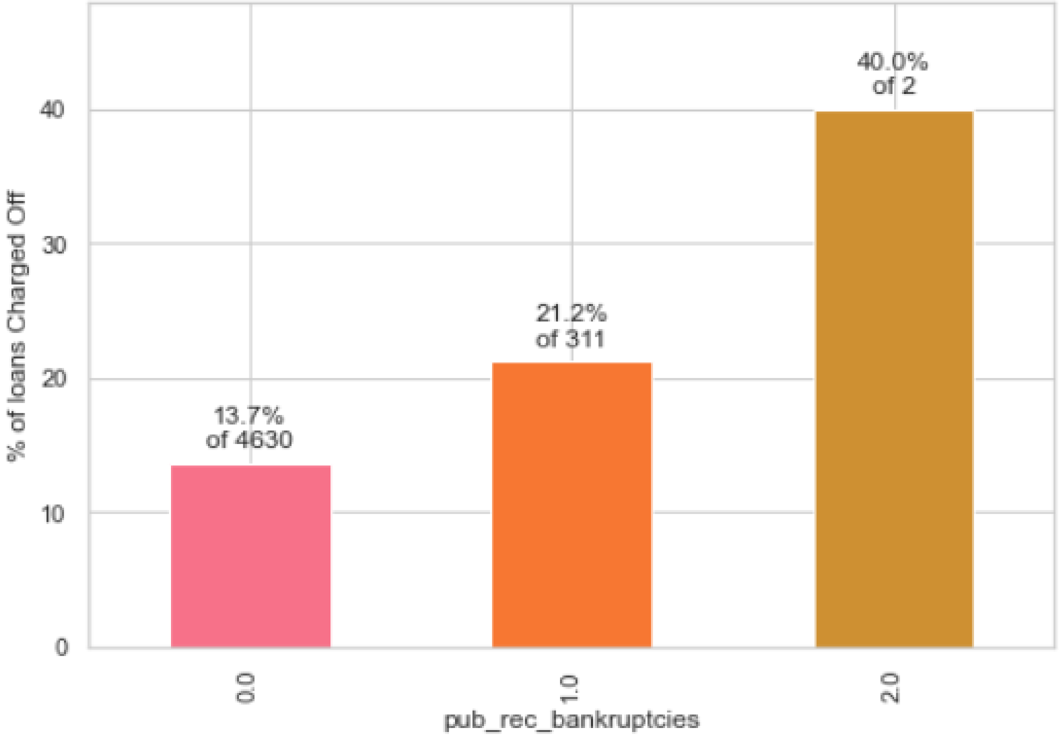
Verified status - Analysis

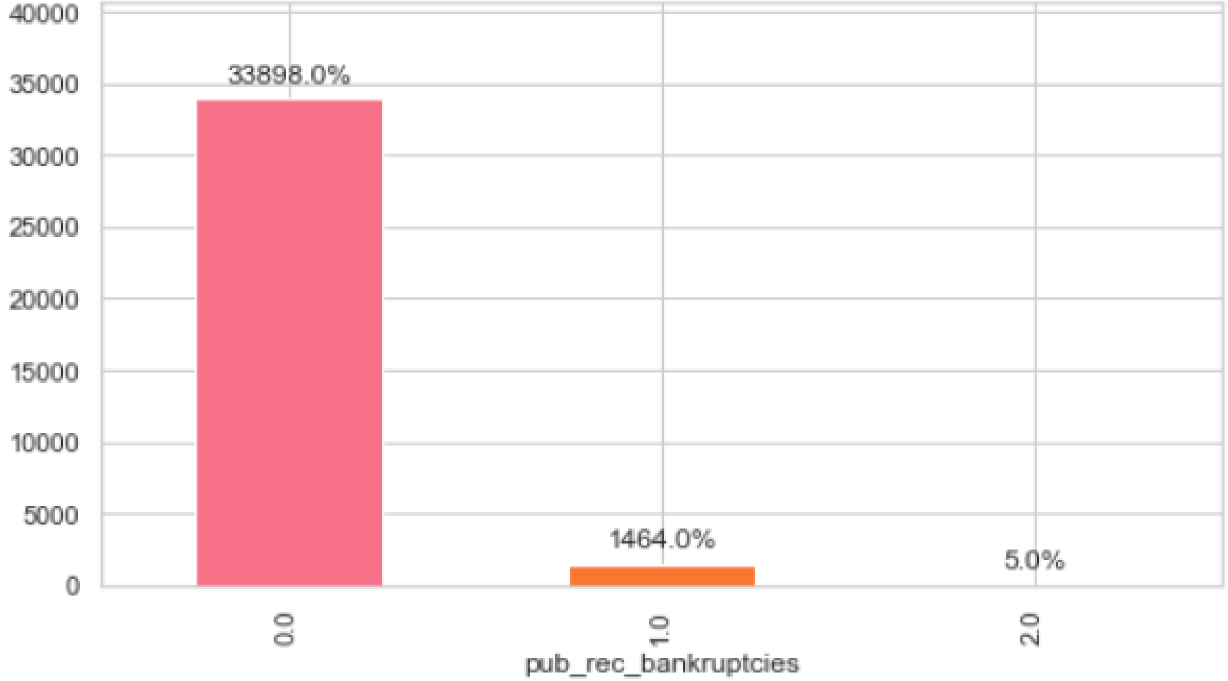


Observation:

- ✓ Overall, it appears that there are more 'Non-Verified' loans than the 'Source Verified' and 'Verified' loans. LC should take steps to verify these loans income to reduce the Charge Off percentage.
- ✓ Verified loans show more charged Off percentage as compared to Not verified loans. This is unexpected and will be investigated later in bivariate analysis

Bankruptcy- Analysis

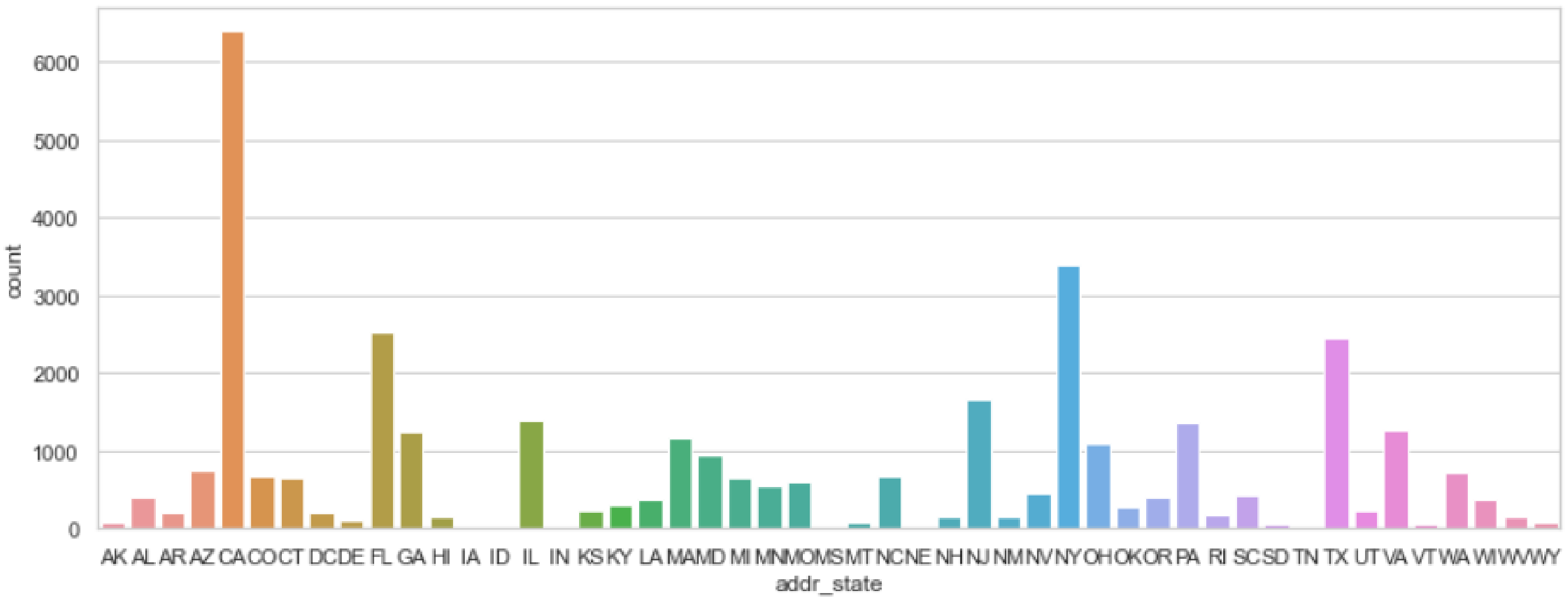


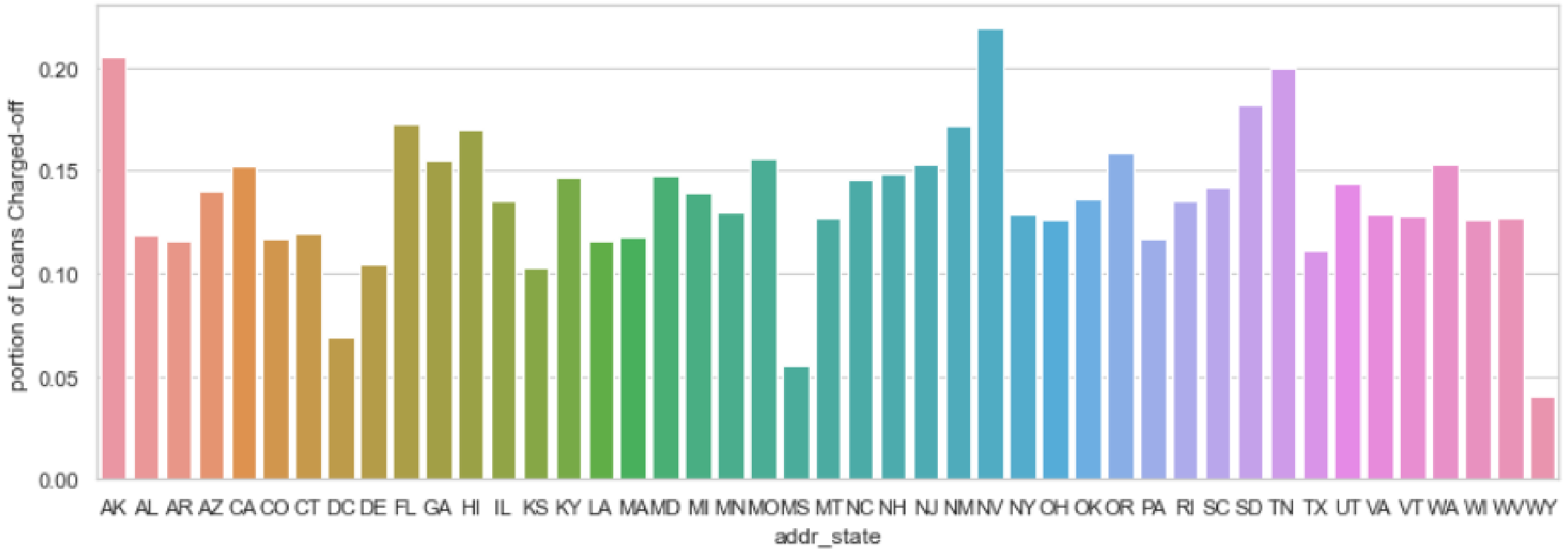


Observation:

- ✓ The percentage of Charged Off loans is markedly higher when the borrower has a prior record of bankruptcy.
- ✓ Very few loans are extended to people with prior record of bankruptcy

Location- Analysis

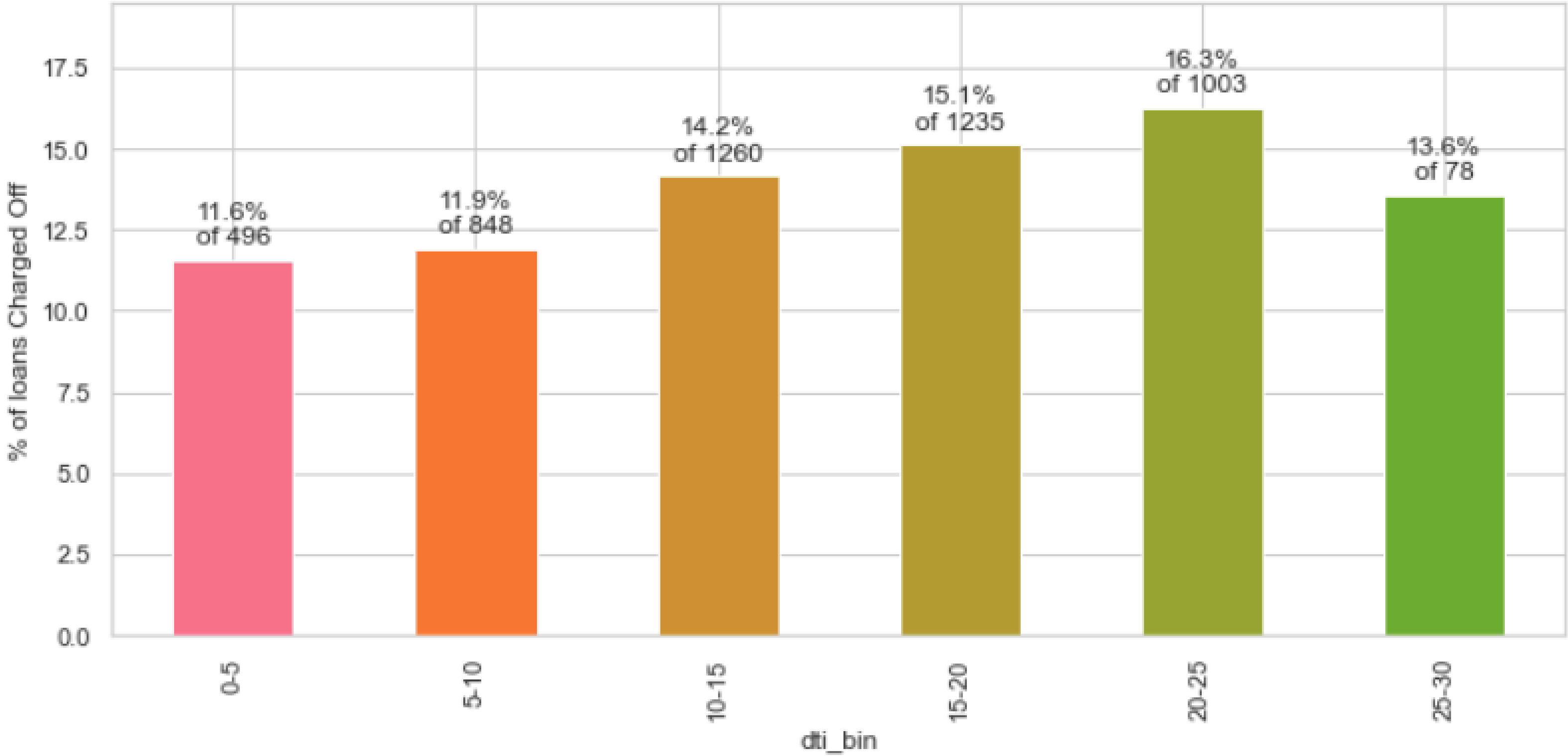




Observation:

- ✓ 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.

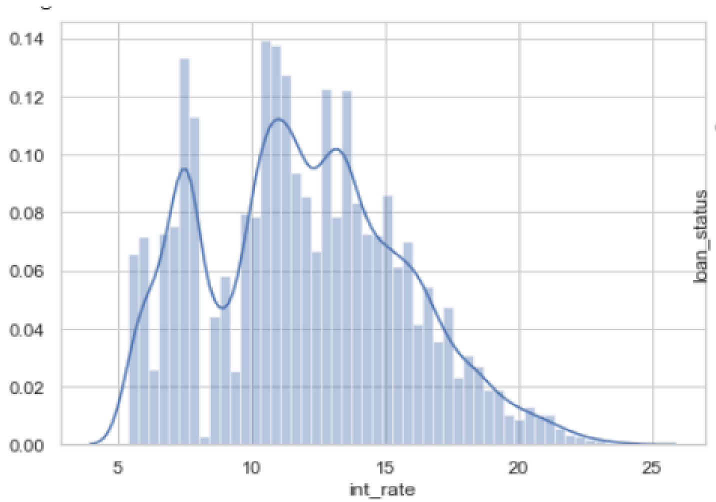
DTI bin- Analysis

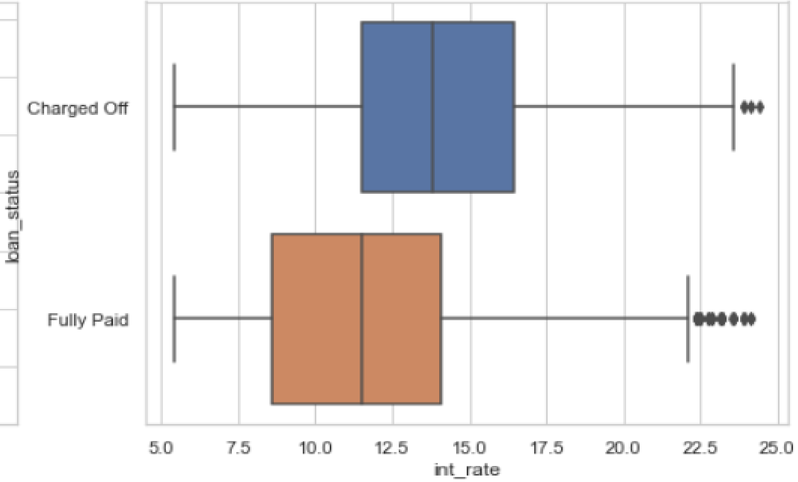


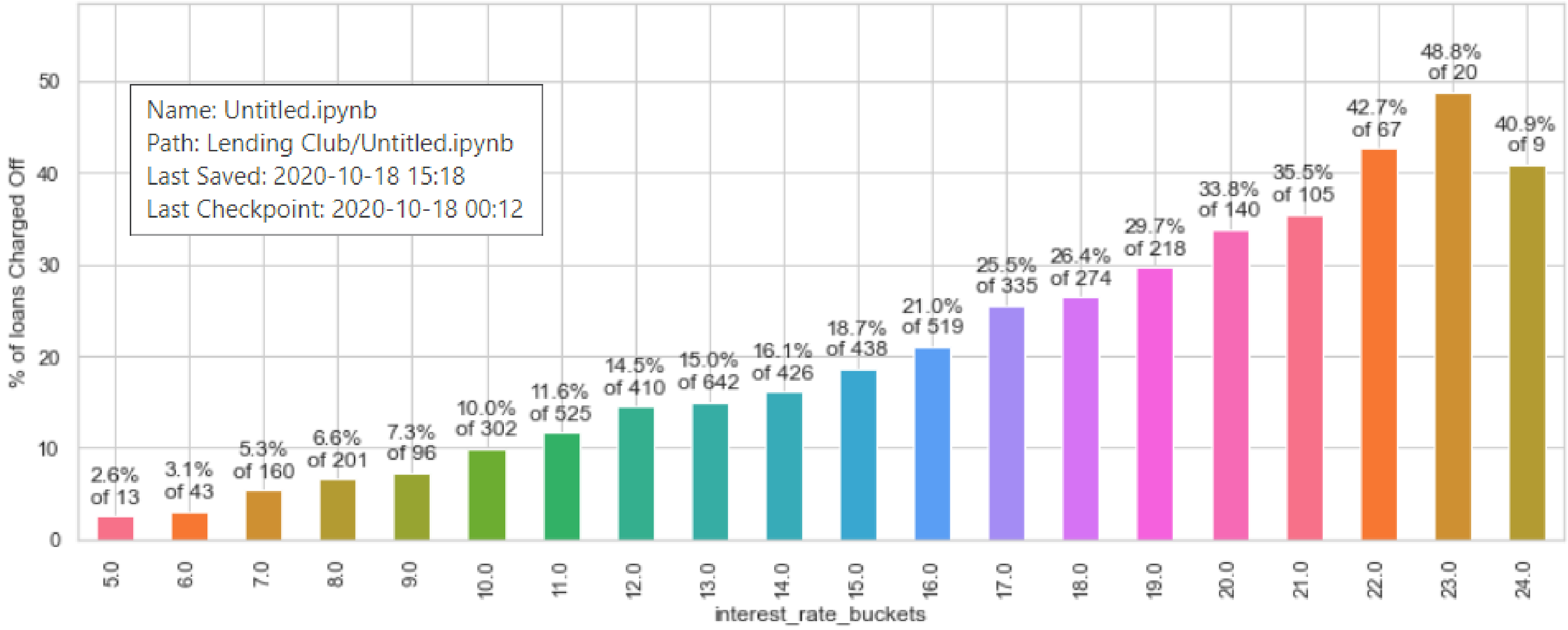
Observation:

- ✓ When the dti or dept payment to income ratio is higher than 20, higher percentage of loans are Charged Off
- ✓ Higher the dti higher the chances of loan being Charged Off.

Interest Rate Analysis





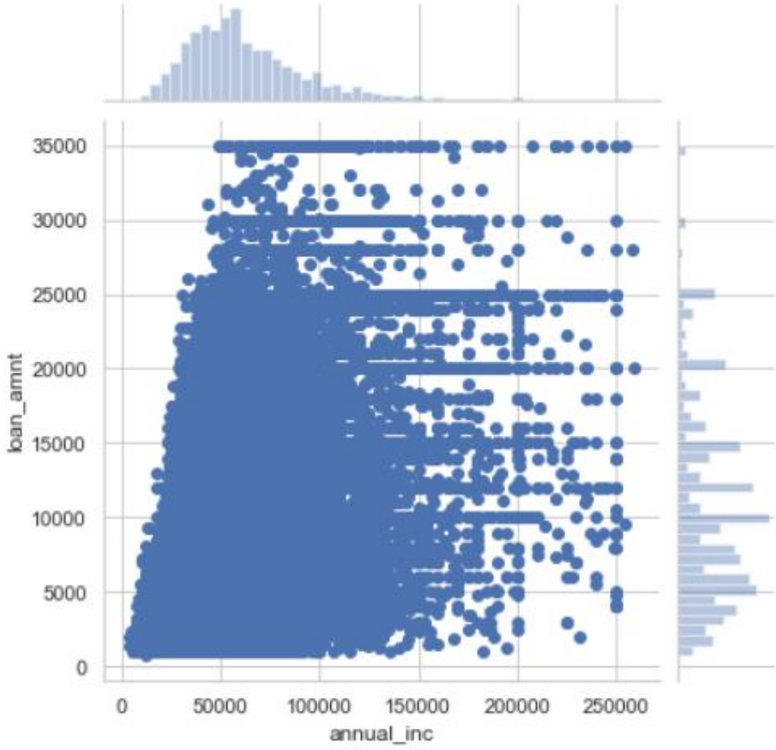


Observation:

- Overall, the interest rate varies from 5.42% to 24.4% with average interest rate of 11.8%.
- The interest rate for Charged Off loans appear to be higher than for Fully paid..
- Percentage of Defaults increases monotonically with higher interest rates. At rates of 19% and above, more than 33% of loans are Charged Off.

Bivariate analysis - Summary

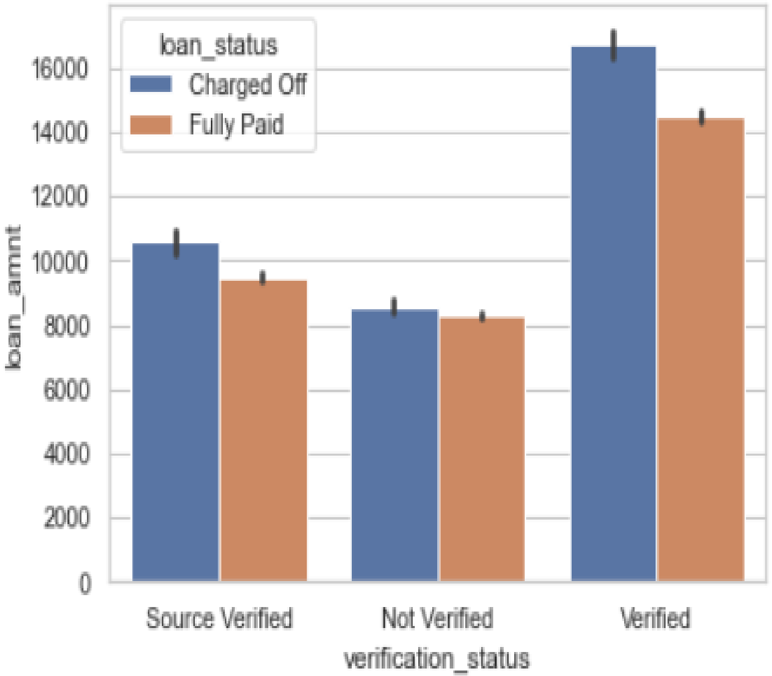
Loan amount vs annual income



Observation:

- There are people with average income lower than 50000 taking loans of 25000 or higher. These would be risky loans.

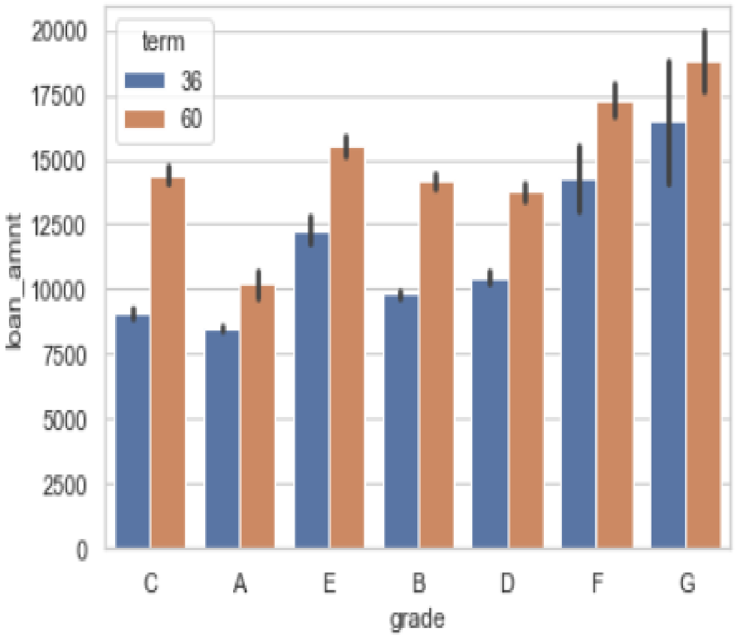
Loan amount vs verification status vs loan status



Observation:

- Higher loan amounts are riskier and are also verified more often by Lending Club

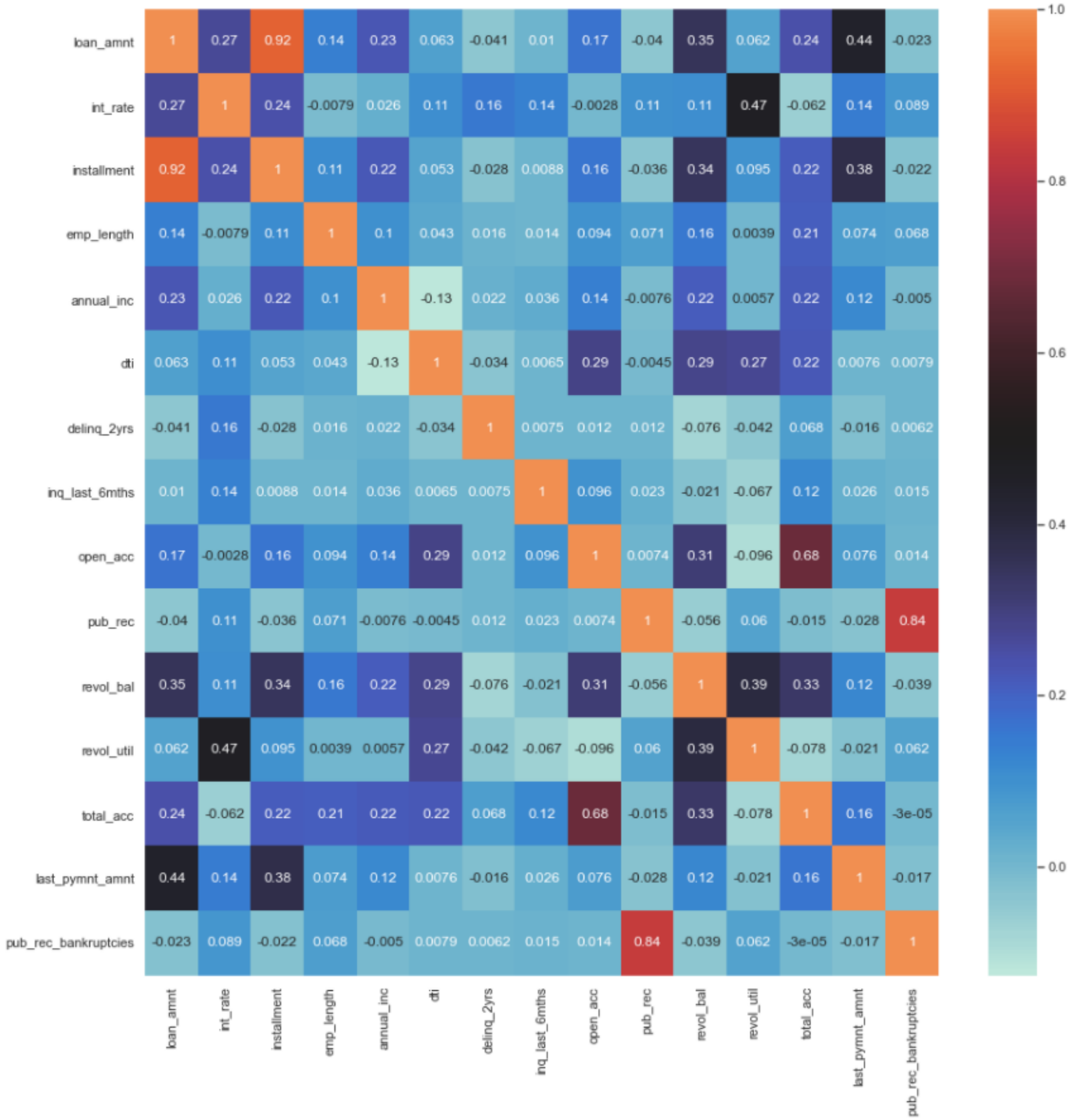
Loan amount vs Grade vs Term



Observation:

- Higher loan amount are associated with lower grade for longer terms.

Bivariate analysis – Revolving utilization vs Grade



Observation:

- Loan amount is correlated to last payment amount with r factor .44, as expected
- Interest rate is correlated to revolving utilization with r factor of .47 - **This is good, as company is charging higher interest from riskier loan.**
- Loan amount and revolving balance 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.

Conclusion

upGrad

Analysis Insight

- ✓ Low grade loans have high tendency to default.
- ✓ Grading system is working as expected.
- ✓ Loans having higher interest rate have more defaulters.
- ✓ Check the background of applicant thoroughly if interest rate is high.
- ✓ Extra scrutiny must be done for the applicants belonging to CA state, as tendency to default is high.
- ✓ When the purpose is debt consolidation check applicant thoroughly as it has high tendency to default

Recommendation

- ✓ Decline loans where mount/income is higher than 30%
- ✓ Minimize number of approvals where purpose is small business
- ✓ Decline high-value loans when revolving line utilization rate greater than 75%
- ✓ Decline 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

Thank You