LENDING CLUB CASE STUDY

Group member:

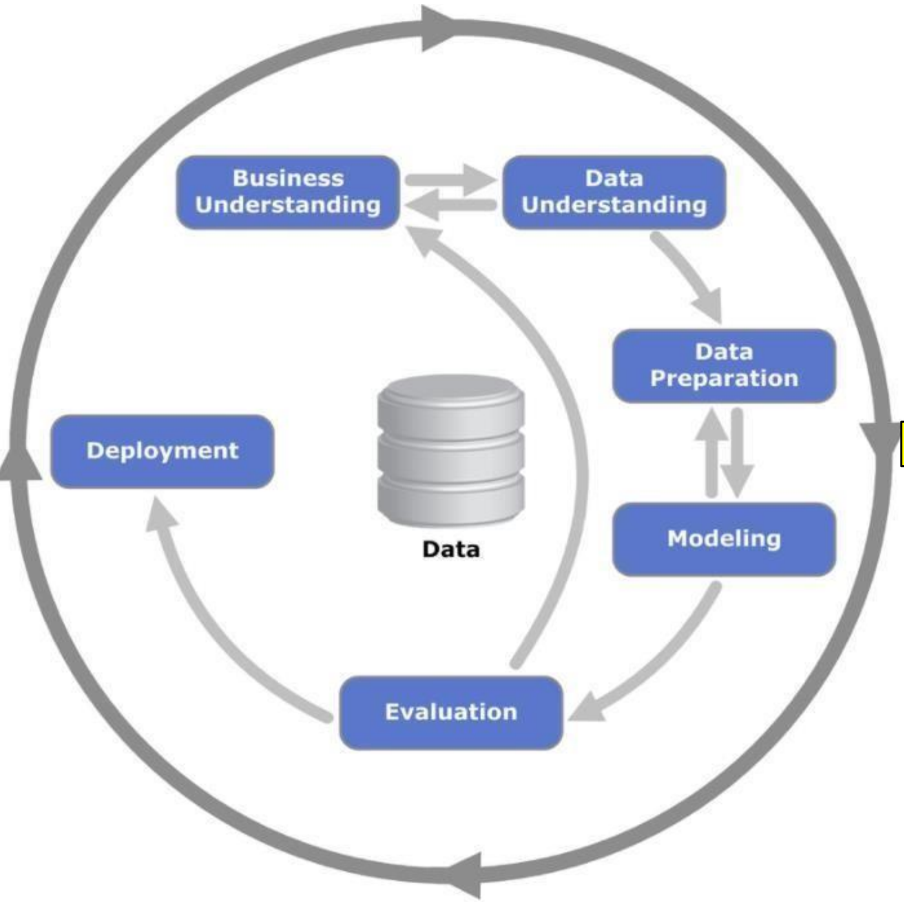
Pushpendra Kumar

Jatinder Kaur

Case Study Objectives

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 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 *consumer attributes* & *loan attributes.* Identification of Loan Applicant traits that tend to 'default' paying back

Problem solving methodology



Understanding Dataset

(Data Cleaning)

Data Cleaning

(Imputed the NA values for all the driver variables)

Data Preparation

Exploratory Data Analysis

(Performed Univariate, Univariate Segmented & Multivariate analysis)

Data Cleaning Steps

- **Delete columns**: Delete unnecessary columns.
- **Remove outliers**: Remove high and low values that would disproportionately affect the results of your analysis.
- **Missing values:** Treat missing values with appropriate approach.
- **Duplicate data:** Remove identical rows, remove rows where some columns are identical.
- **Filter rows:** Filter by segment, filter by date period to get only the rows relevant to the analysis.

Analysis

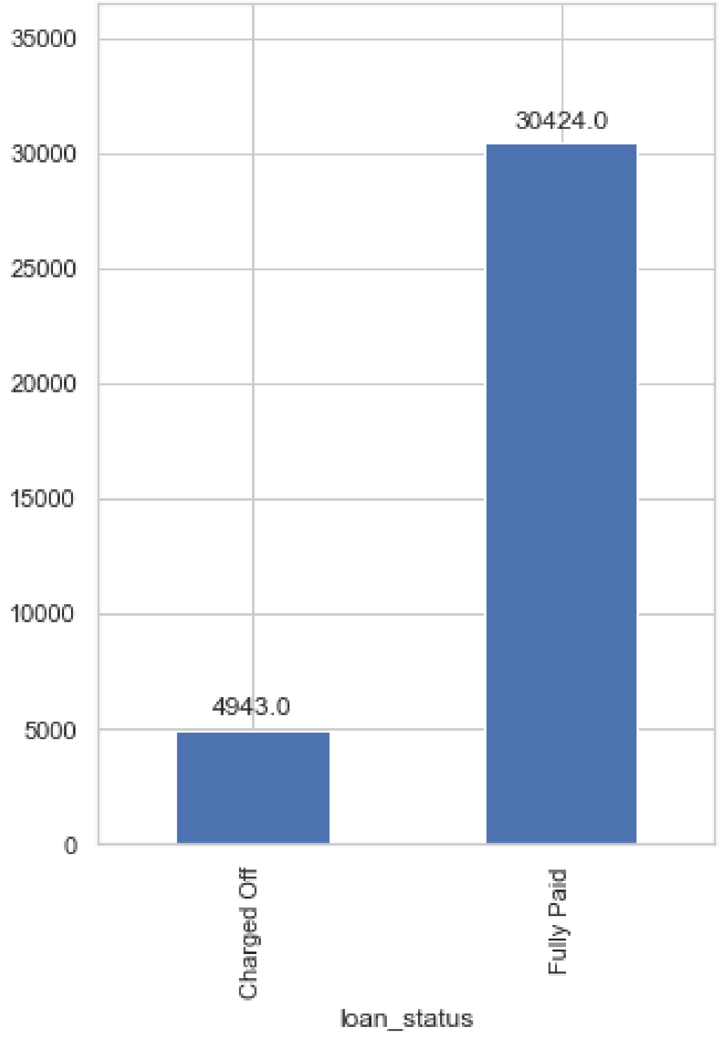
- The essence of the whole project is to analyze and understand how consumer attributes and loan attributes are influencing the tendency of defaulting.
- We performed **data cleaning and preparation** on the Loan dataset:
  - Imputed the NA values for all the variables
  - Created two new columns:
    - Profit and Loss column
    - Ratio of funded amount and annual income
- During **univariate analysis** we have created:
  - Histograms and Bar charts to check out the distribution of all the driver variables
  - Box plots to detect the Outliers
  - Performed the Multivariate analysis to understand how different variables interact with each other.

Loan Status Analysis

Total Loans

Approximately 14% of loans are defaulted

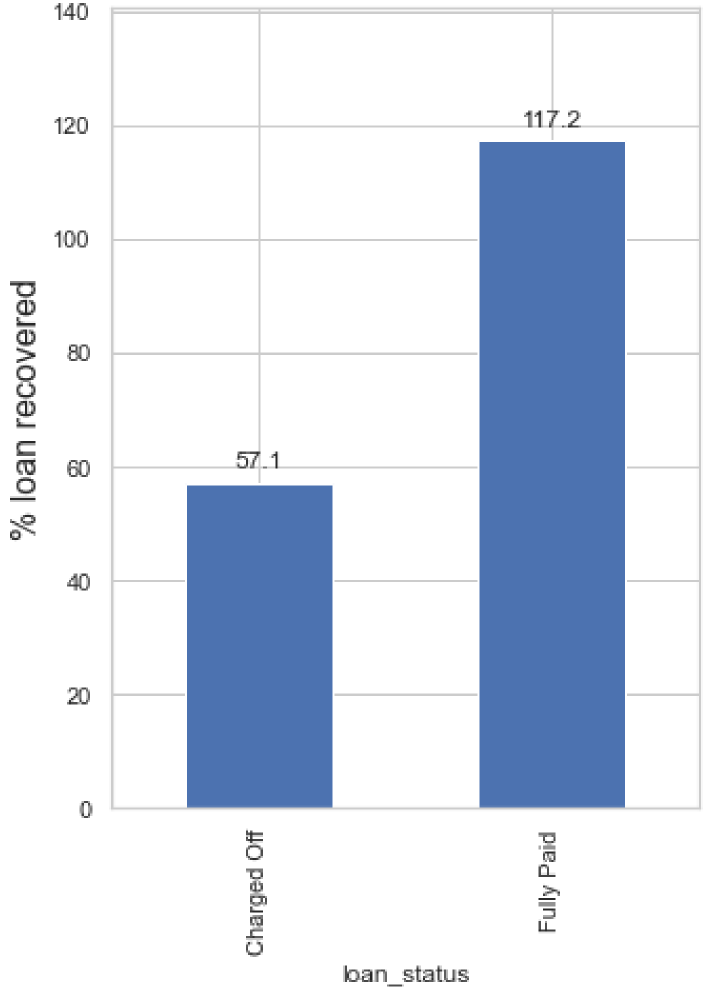
Any variable that increases percentage of default to higher than 16.5% should be considered a business risk.



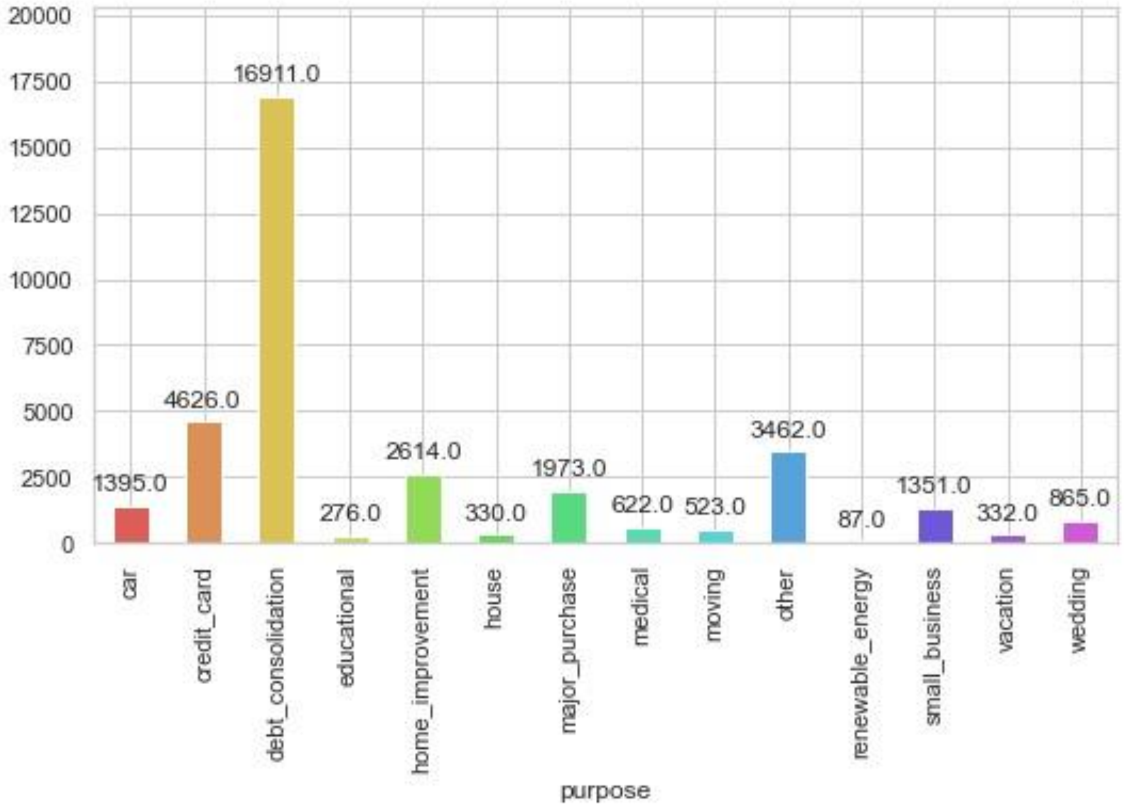
Total Money Earned

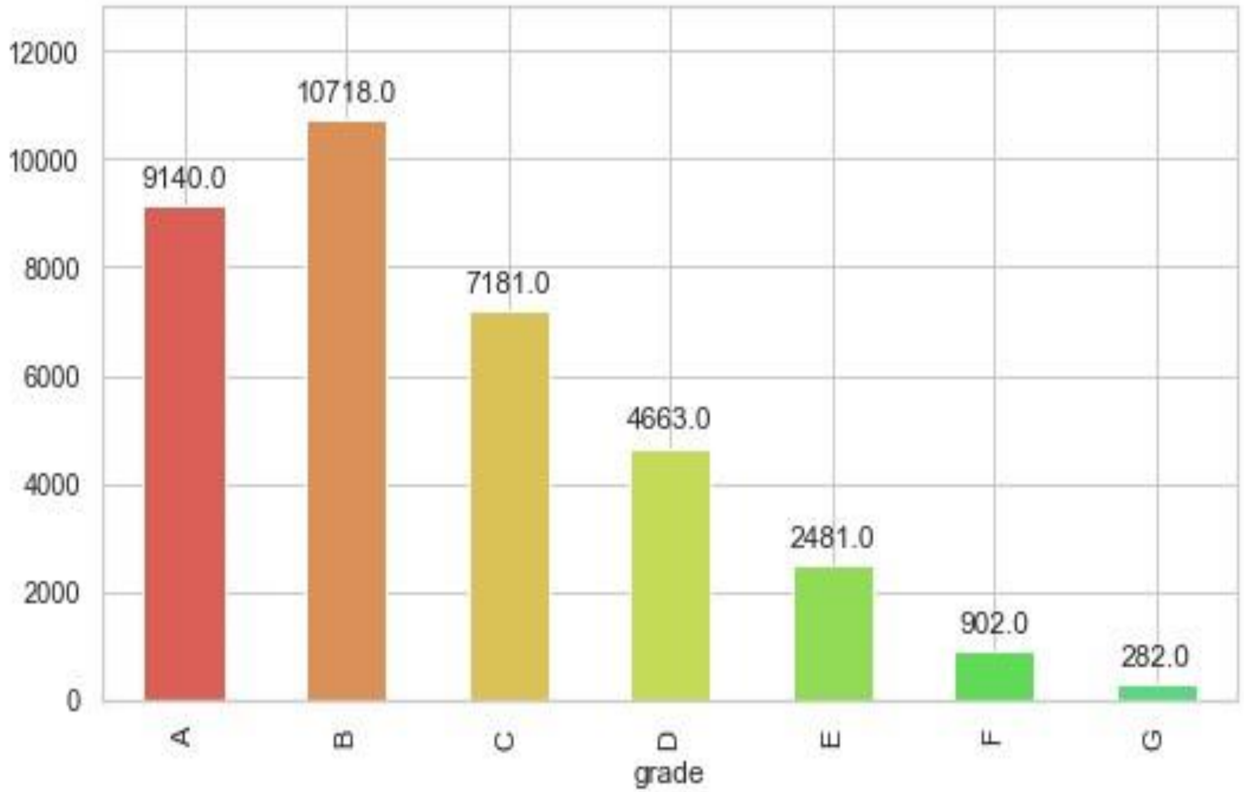
Lending Club only recovers 57% of the loan amount when loans are **defaulted**.

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



Loan Grade & Purpose Analysis



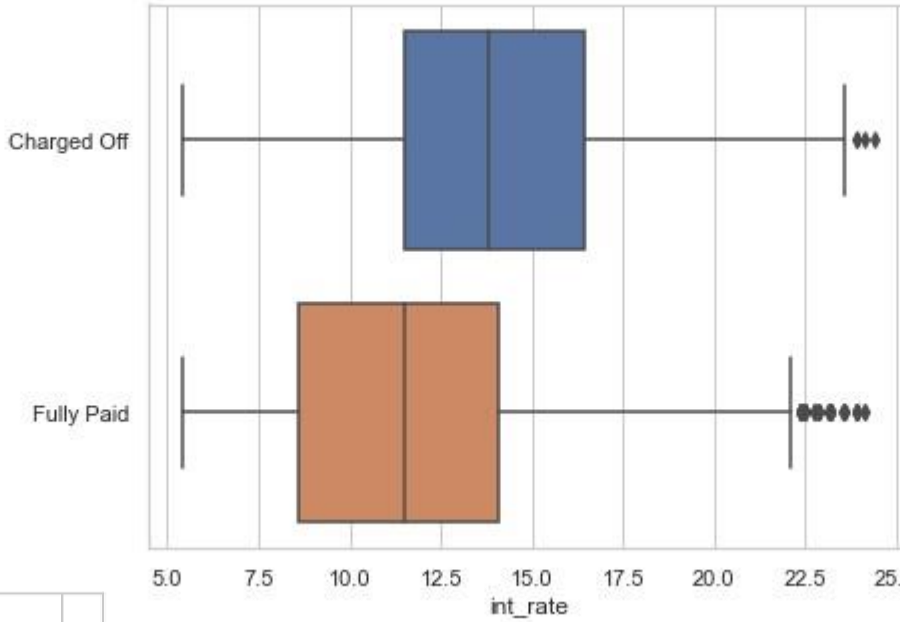


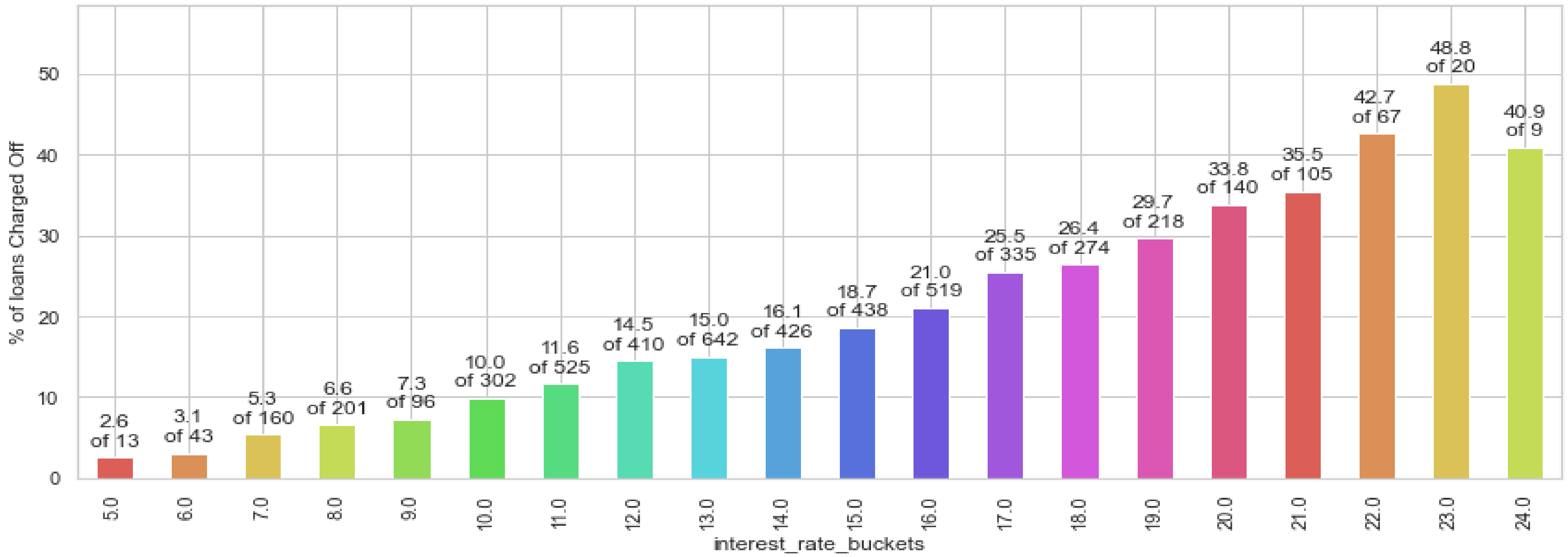
Maximum number of loans are for **debt consolidation**, followed by credit card

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

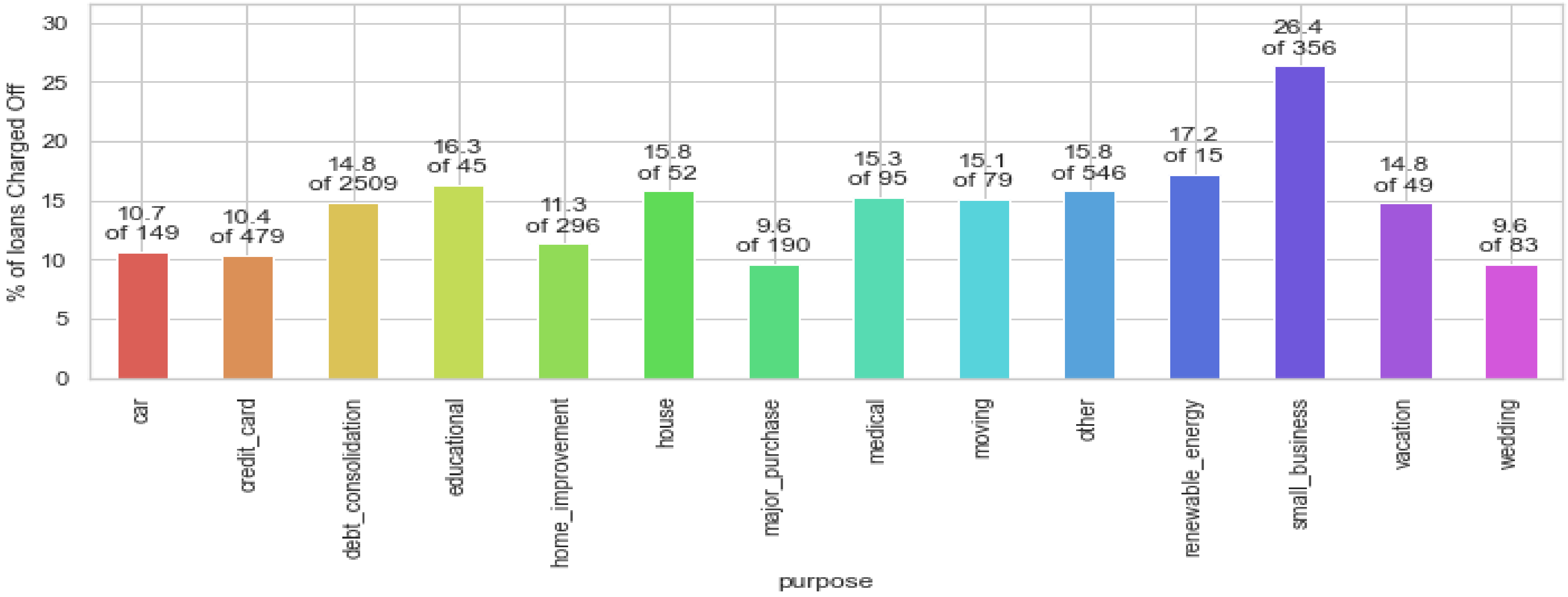
Analysis - Defaults by Interest Rate

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



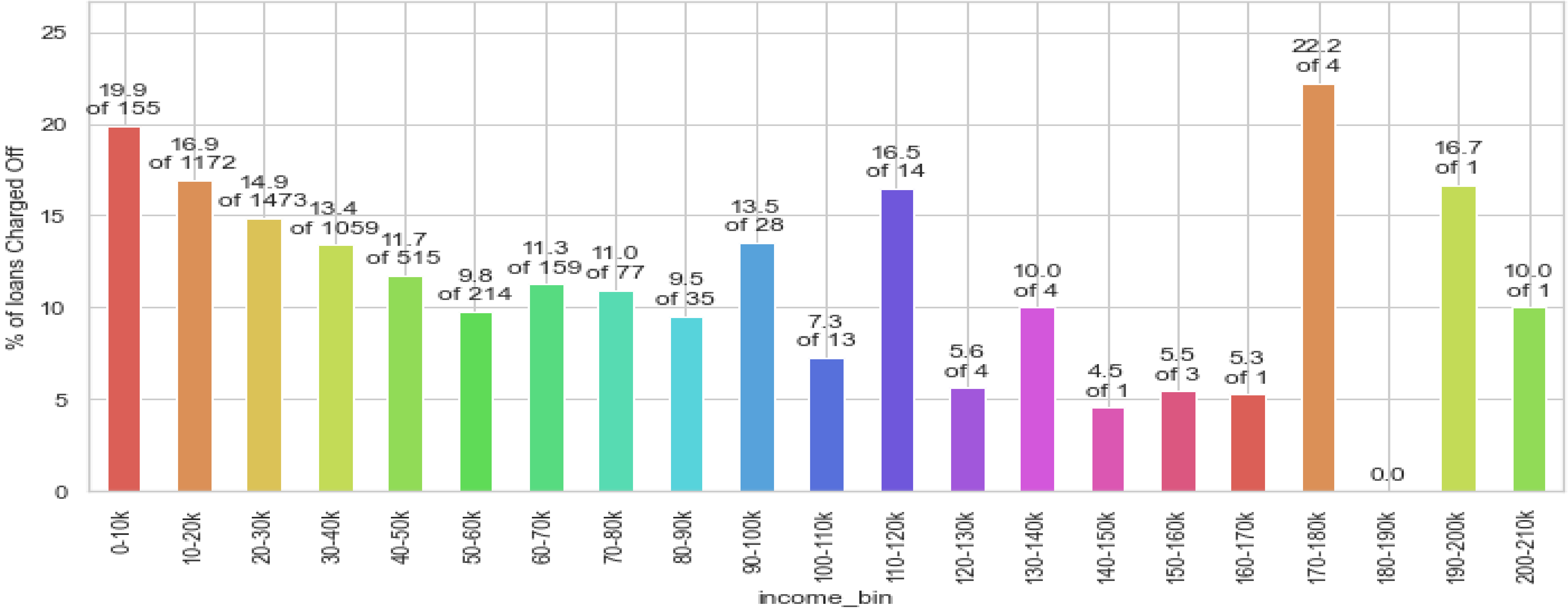


Loan Amount Analysis by Loan Purpose



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

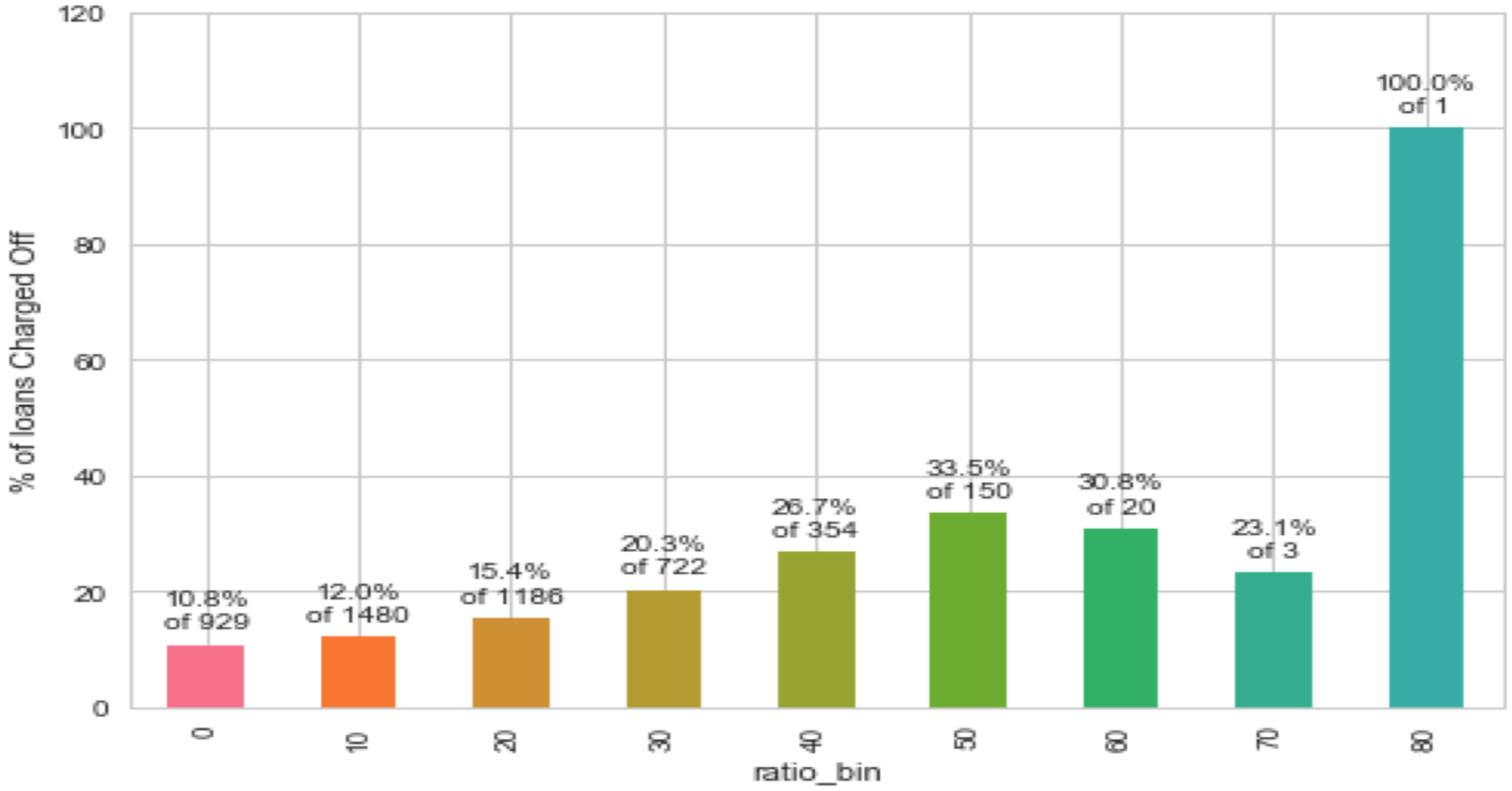
Income Analysis for the loan applicants



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

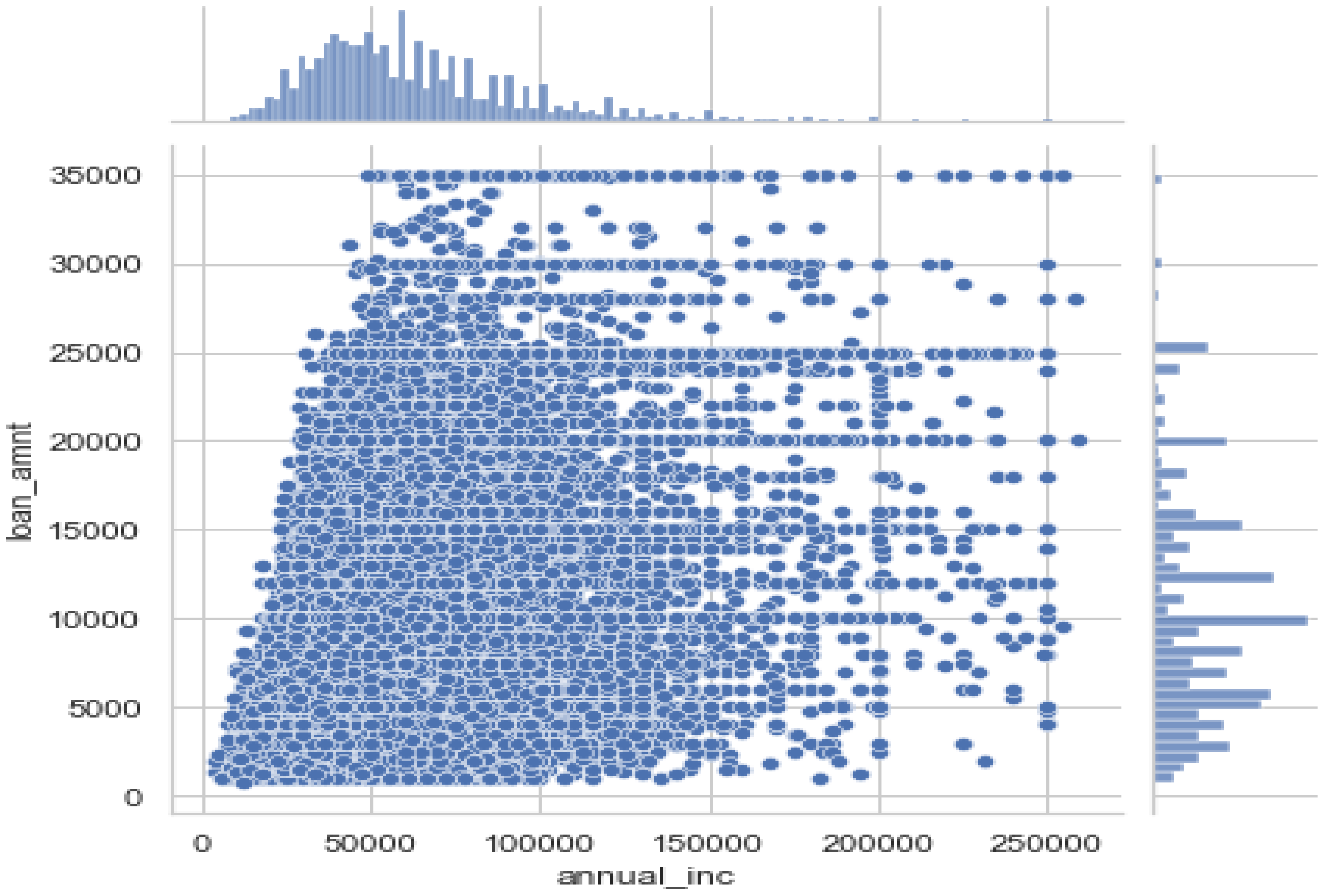


As long as loan amount is less than 20% of annual income, defaults are low.

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

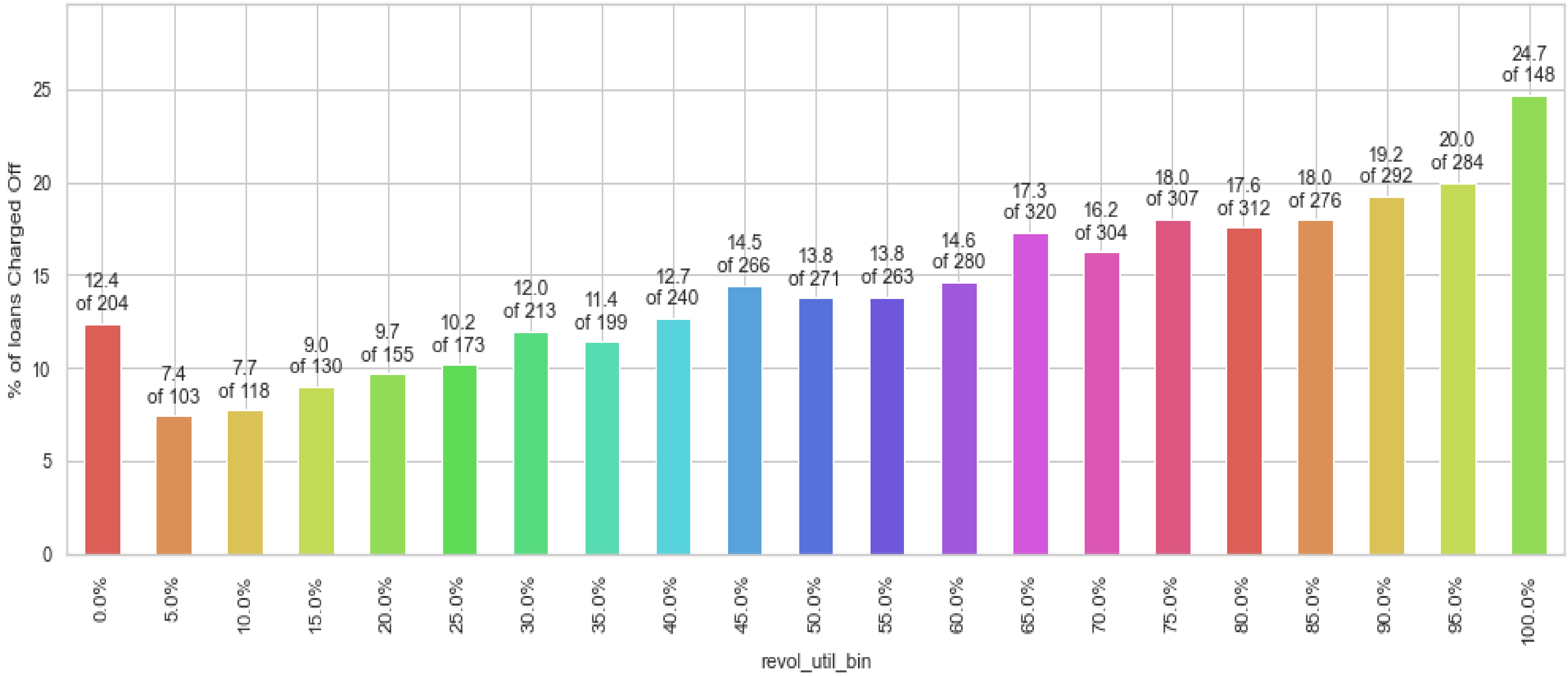
Analysis - Defaults by ratio Continued

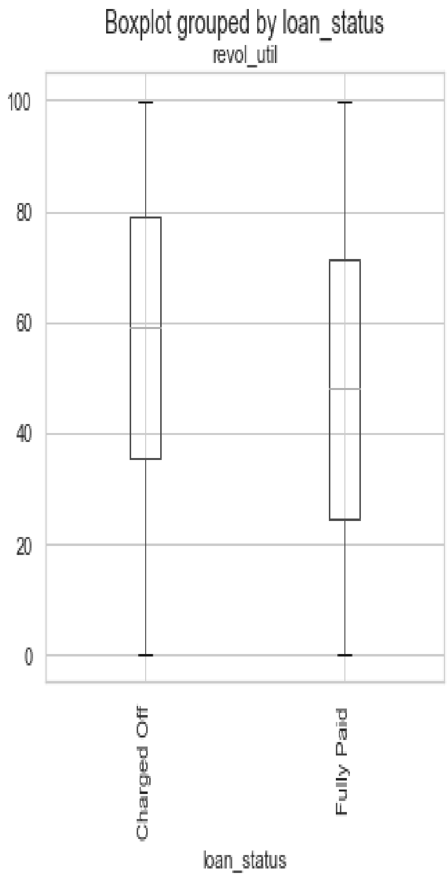


Here Lending Club has extended high-value loans to people with low income.

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

Analysis - Defaults by Revolving Line Util Rate

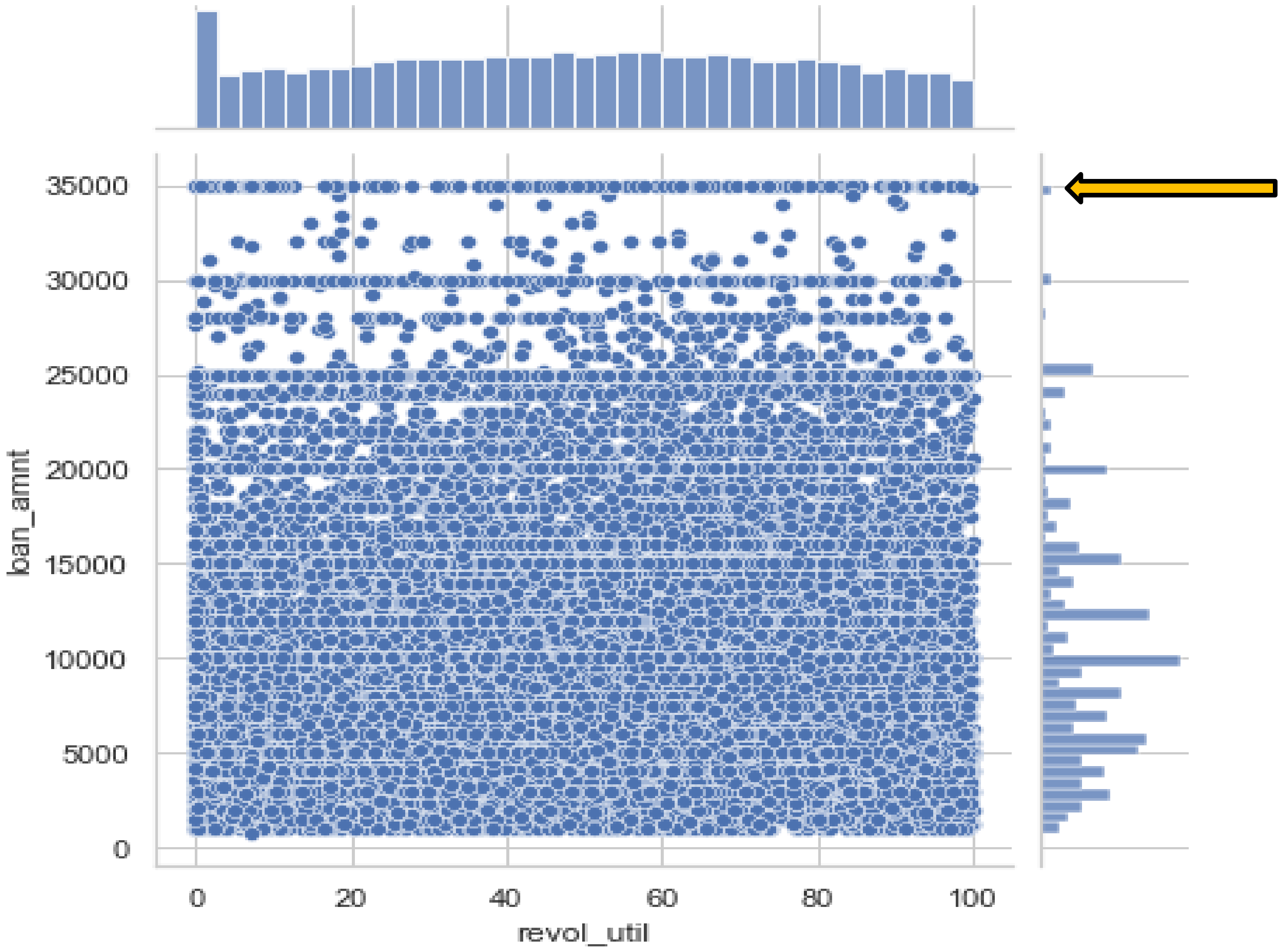




People with high utilization of Revolving Line of Credit at the time of taking loan default more.

Loans with utilization > 75% are risky.

Analysis - Defaults by Rev Util Continued

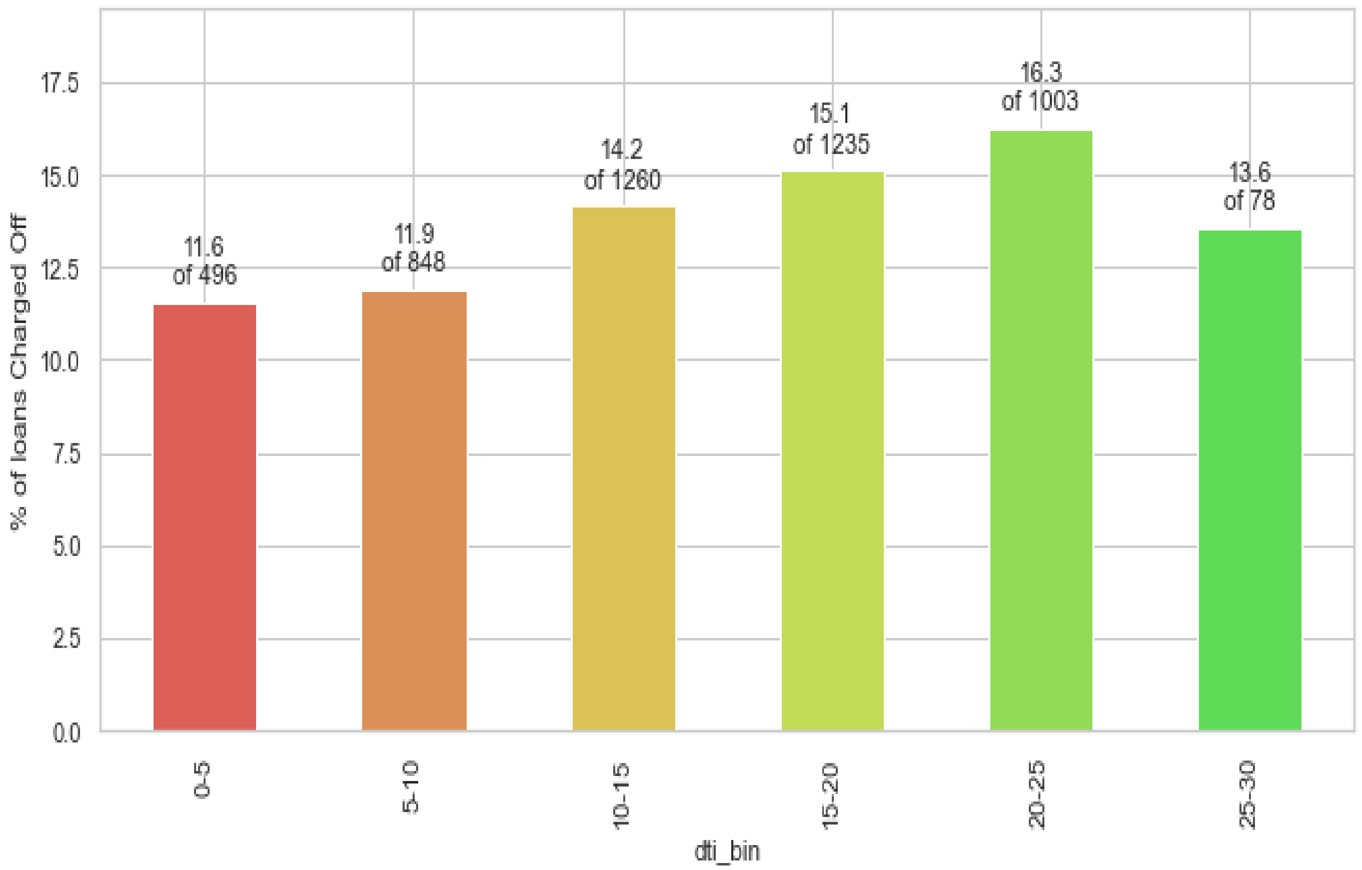


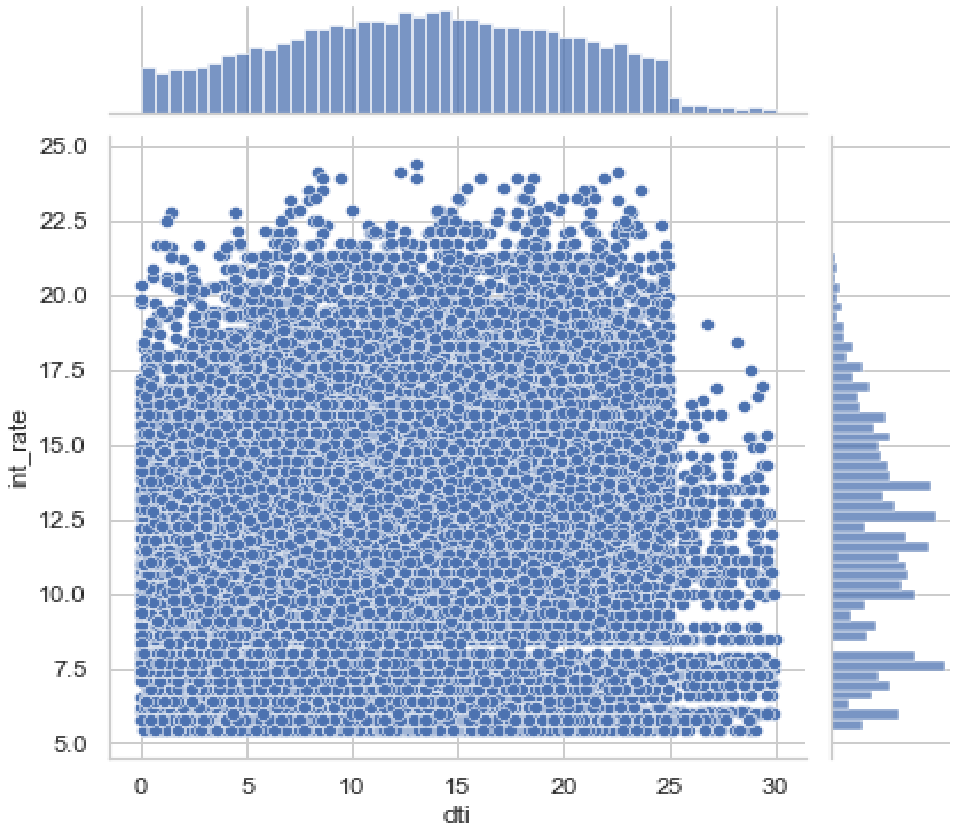
There have been some high value loans extended to borrowers with revolving line utilization rate of higher than 75%.

This practice should be stopped.

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

Analysis – Defaults by Debt to Income Ratio

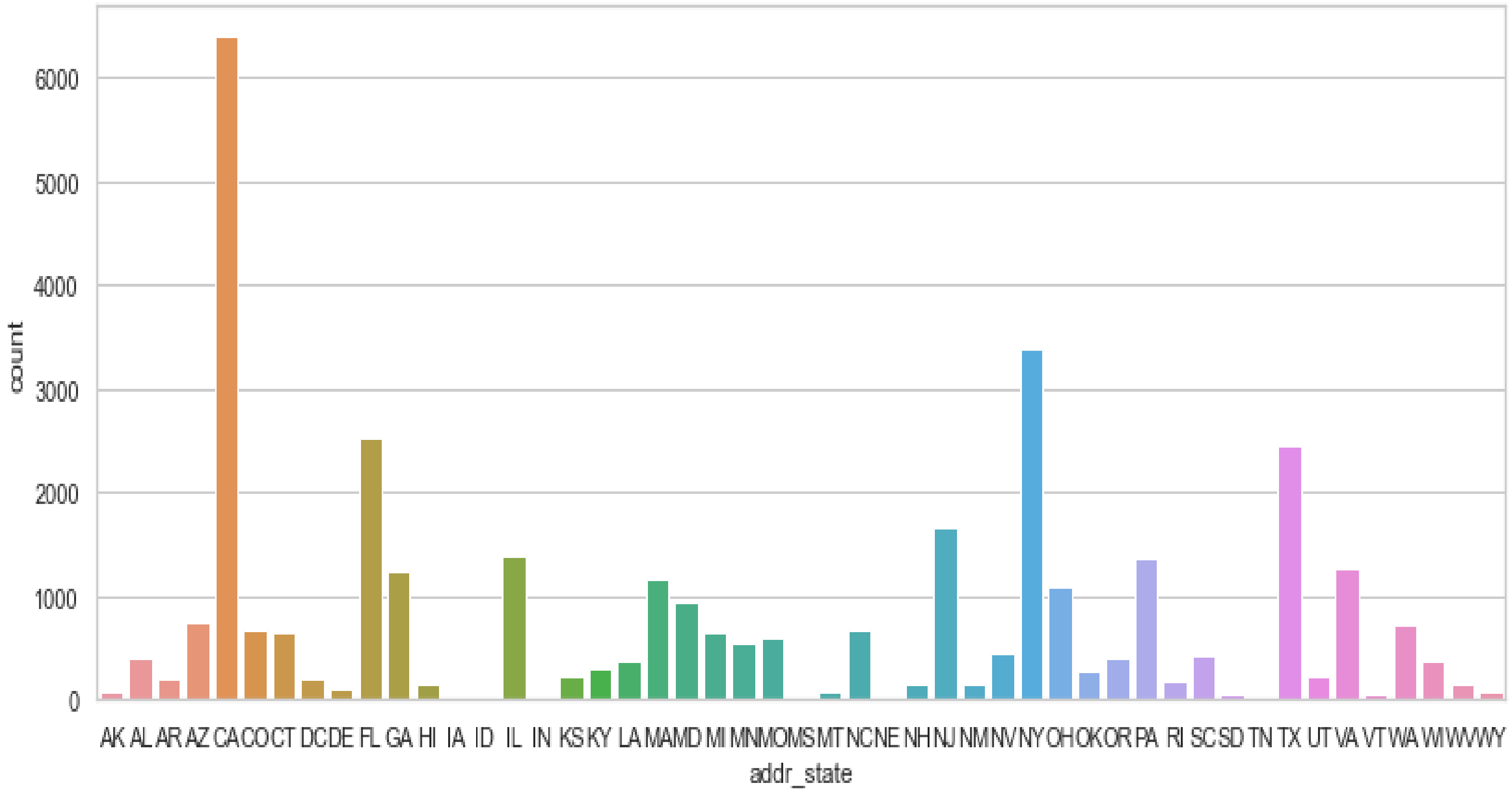


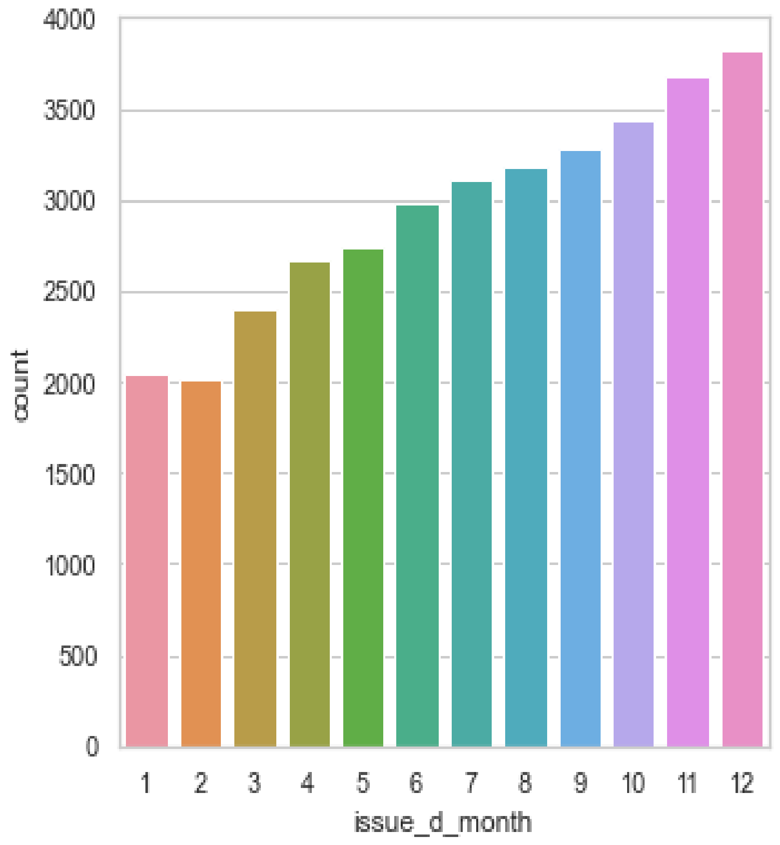


Percentage of default rises with dti ratio. As the dti ratio rises above 20, the loans become risky.

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

Analysis Loan By State & Loan Number

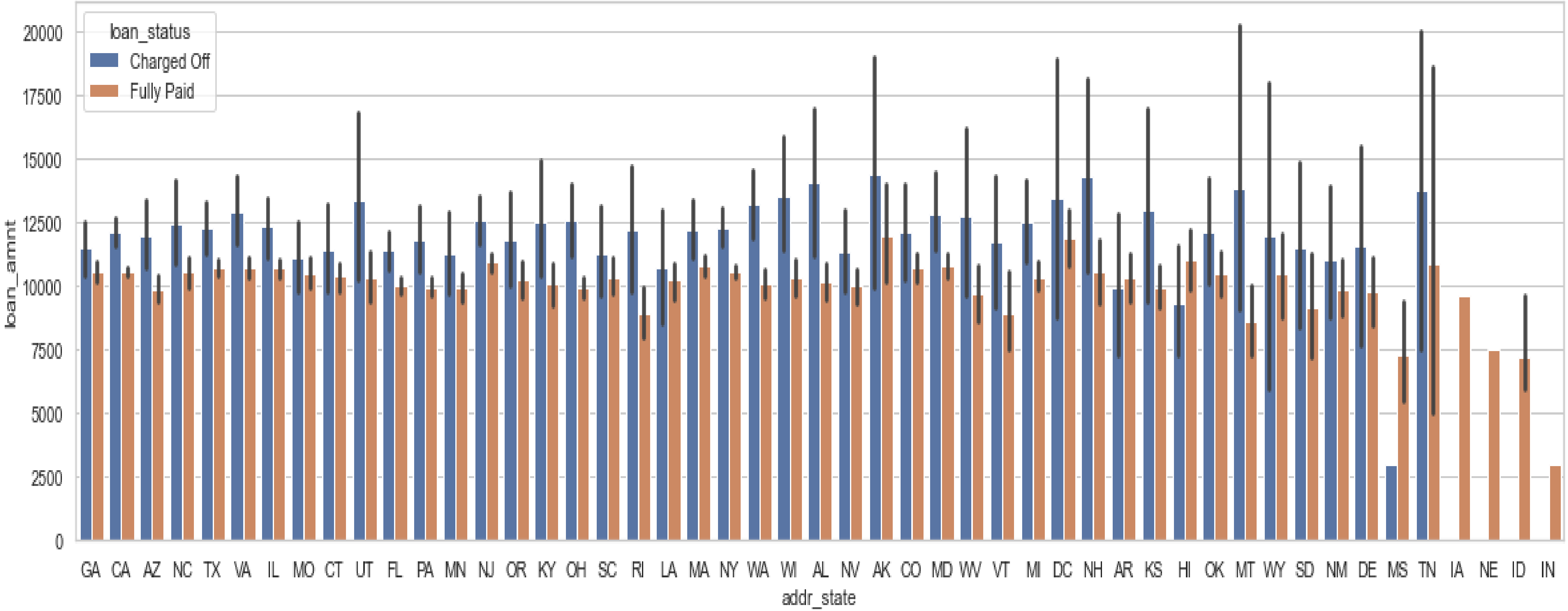




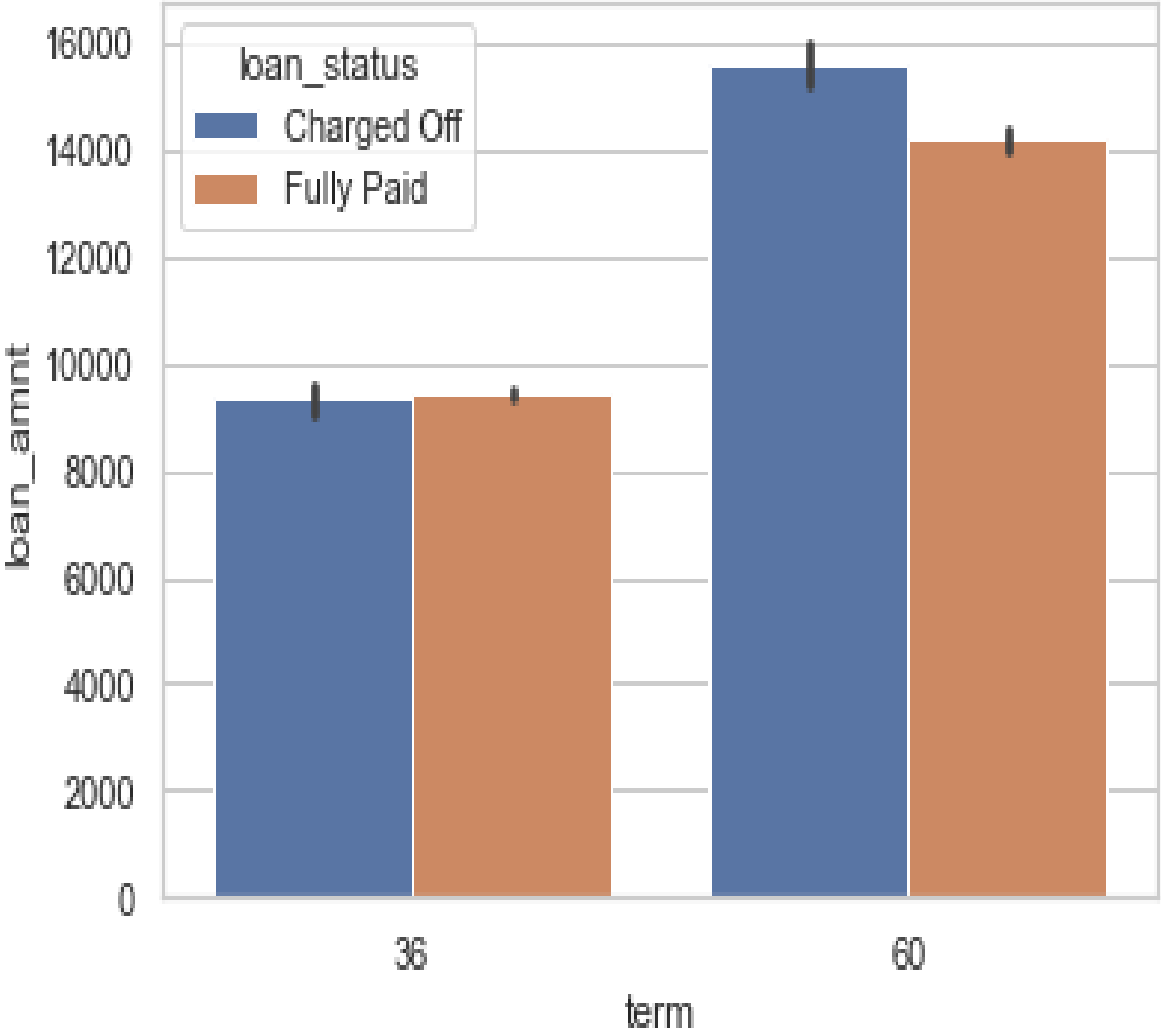
Maximum loans are in populous states, California, New York, Florida and Texas

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

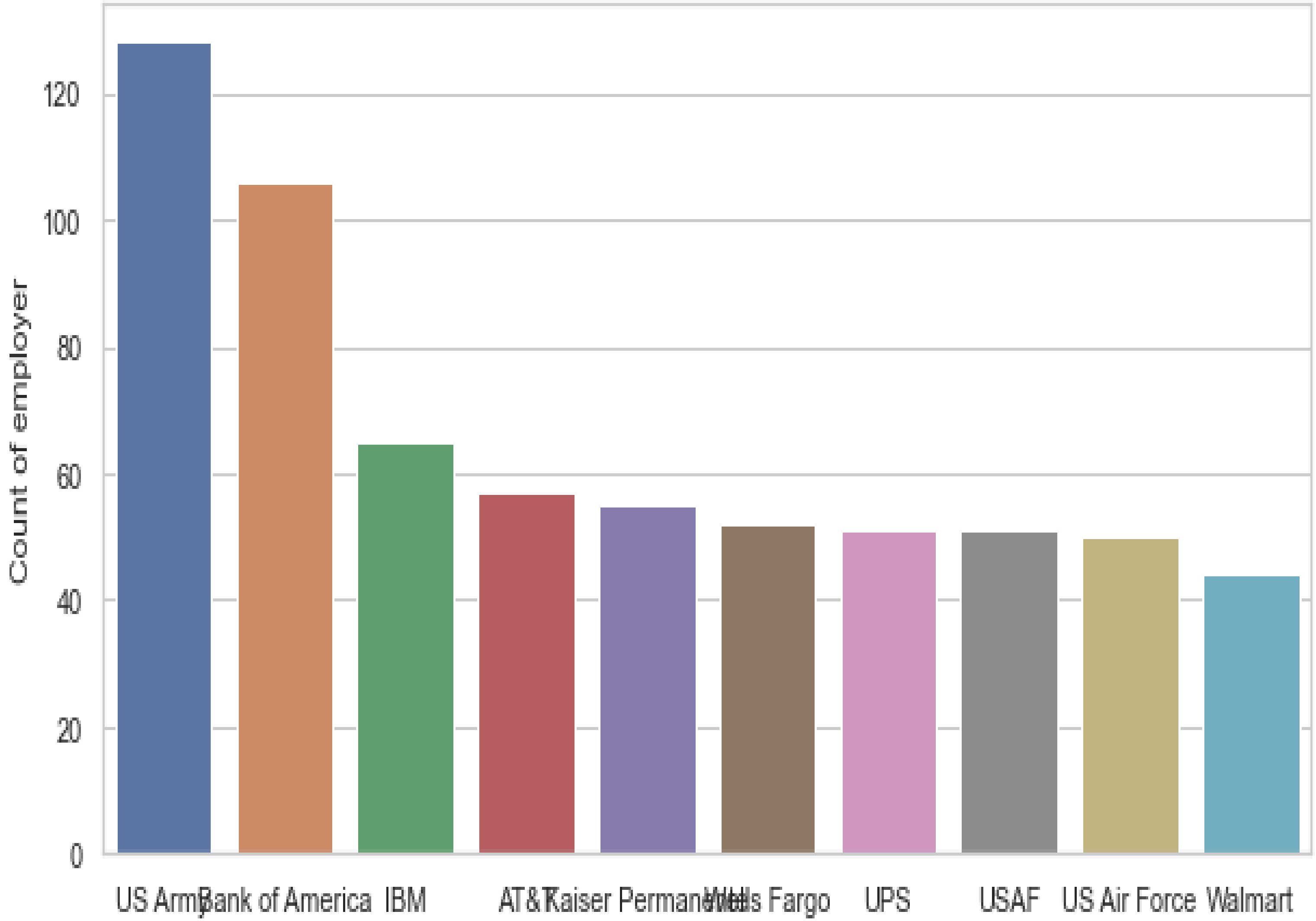
Analysis – State vs. Loan amount vs. Loan status



It shows state wise loan status with loan amount.



It shows maximum loan amount charged off at 60 month term.



- Max of the loan takers are from the US Army, then Bank of America followed by IBM and AT&T.
- Kaiser Permanents, wells Fargo, UPS, USAF, US Air Force and Walmart are also among the top 10 emp_title loans.

Recommendations

Better Quality Borrowers

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

Conclusion

Results

1. Low grade loans have high tendency to default. Grading system is working as expected.

2. Loans having higher interest rate have more defaulters. Check the background of applicant thoroughly if interest rate is high.

3. Extra scrutiny must be done for the applicants belonging to CA state, as tendency to default is high.

4. When the purpose is debt consolidation check applicant thoroughly as it has high tendency to default.