# Gramener Case Study

**Team members:**

Ankur Mathur

Apurv Jain

Udit Sethi

Kumar Avinash

# Problem Statement

This company 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.

Like most other lending companies, lending loans to ‘risky’ applicants is the largest source of financial loss (called credit loss). The credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed. In other words, borrowers who **default** cause the largest amount of loss to the lenders. In this case, the customers labelled as 'charged-off' are the 'defaulters'.

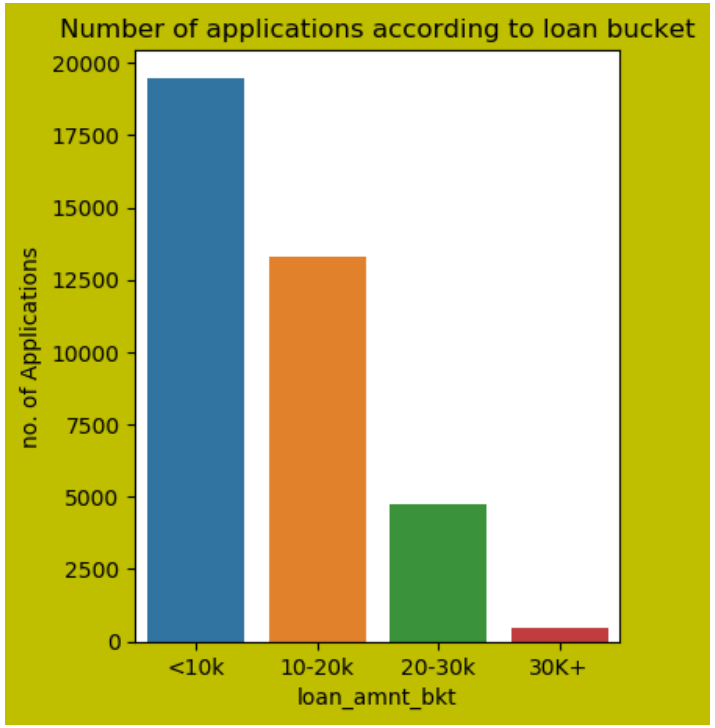
If one is able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. Identification of such applicants using EDA is the aim of this case study.

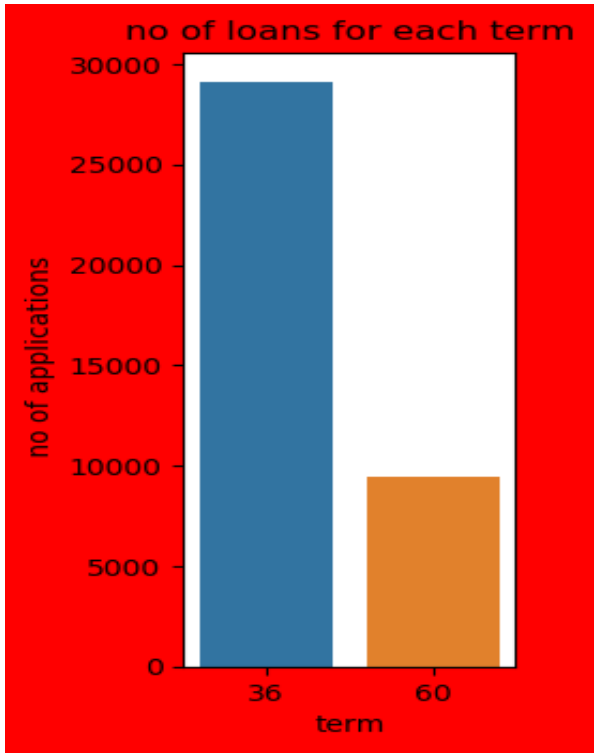
# Analysis approach

- For the given data set, we first need to clean the data for further analysis. The cleaning of data involves removing columns with significant percentage of null values, removing rows with various values being null, changing the datatypes of certain columns for better analysis.
- As the problem statement concerns us with the Defaulters in the given data set, we also remove the current loan holders as they do not add value to the analysis.
- Post this there are buckets made for various parameters for deeper analysis, which is followed by univariate and bivariate analysis, which points us to the trends which can help us identify the triggers which lead to a loan default. This can in turn suggest the loan issuer to practice restraint while disbursing loans.

# Solution

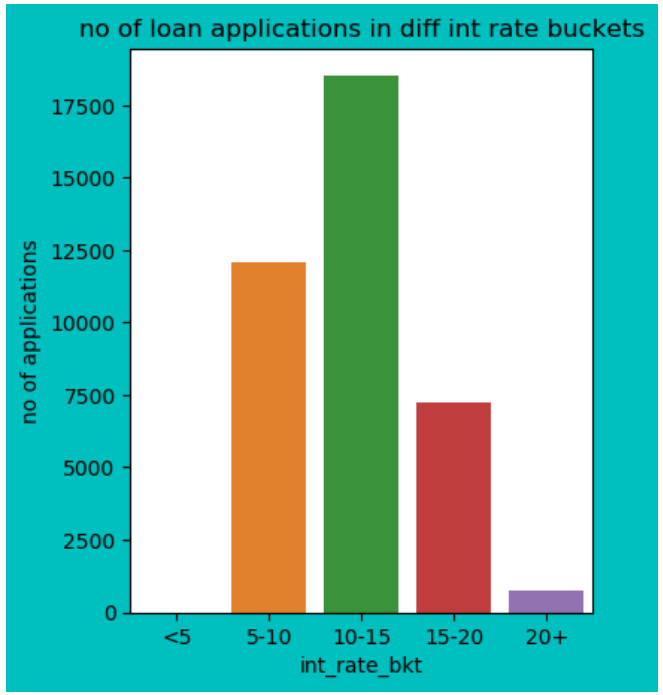
- After the data cleaning is done and we get rid of all the empty columns and rows and then convert the columns to correct data types, we extract the month and date from the issue date and made as separate columns for deeper analysis.
- We then bucket the loan amount, funded amount, funded amount by investors and the interest rates in various intervals. We also create buckets for annual income and debt to income ratios for better understanding of the data.
- Starting with our univariate analysis here, we find that 50% of the loans are for less than 10000, 34.5% for 10K-20K range, 12% for the 20K-30K range and 1.2% for more than 30K.
- Another information which we get from the analysis is that 75.4% of loans were given for a duration of 36 months and remaining for the 60 months period

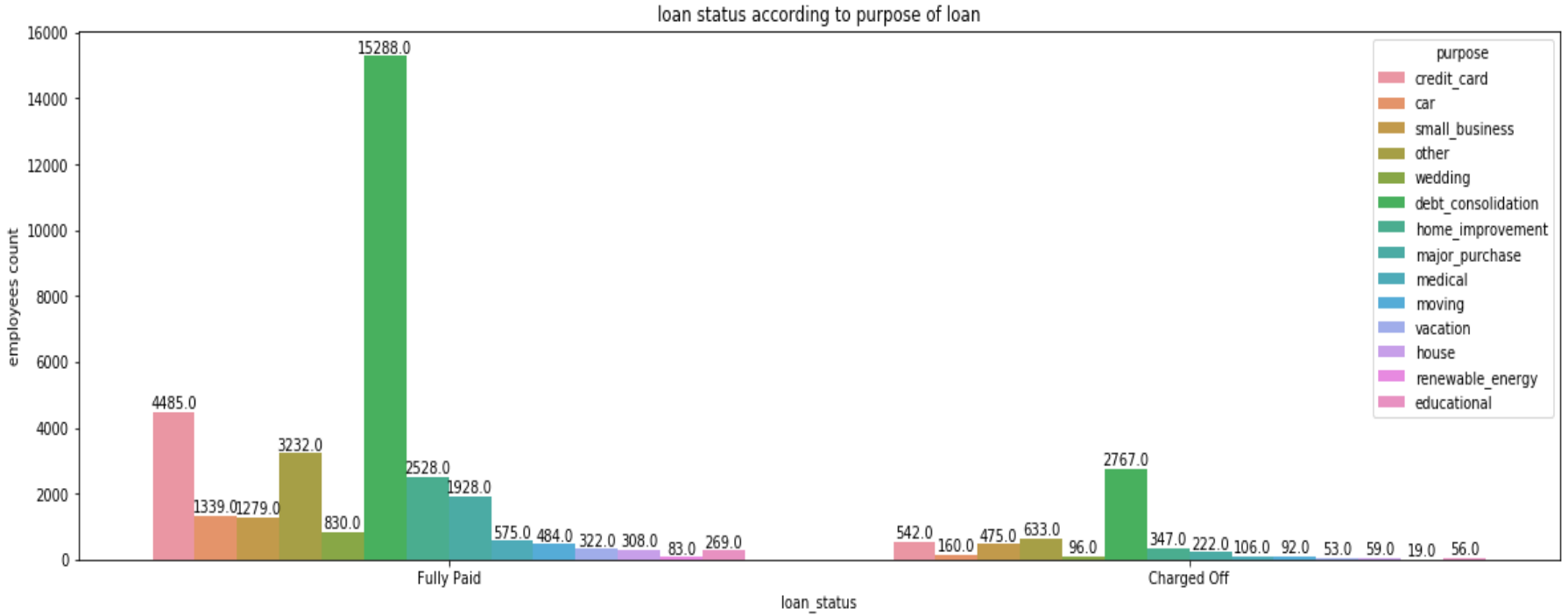


# Solution – contd.

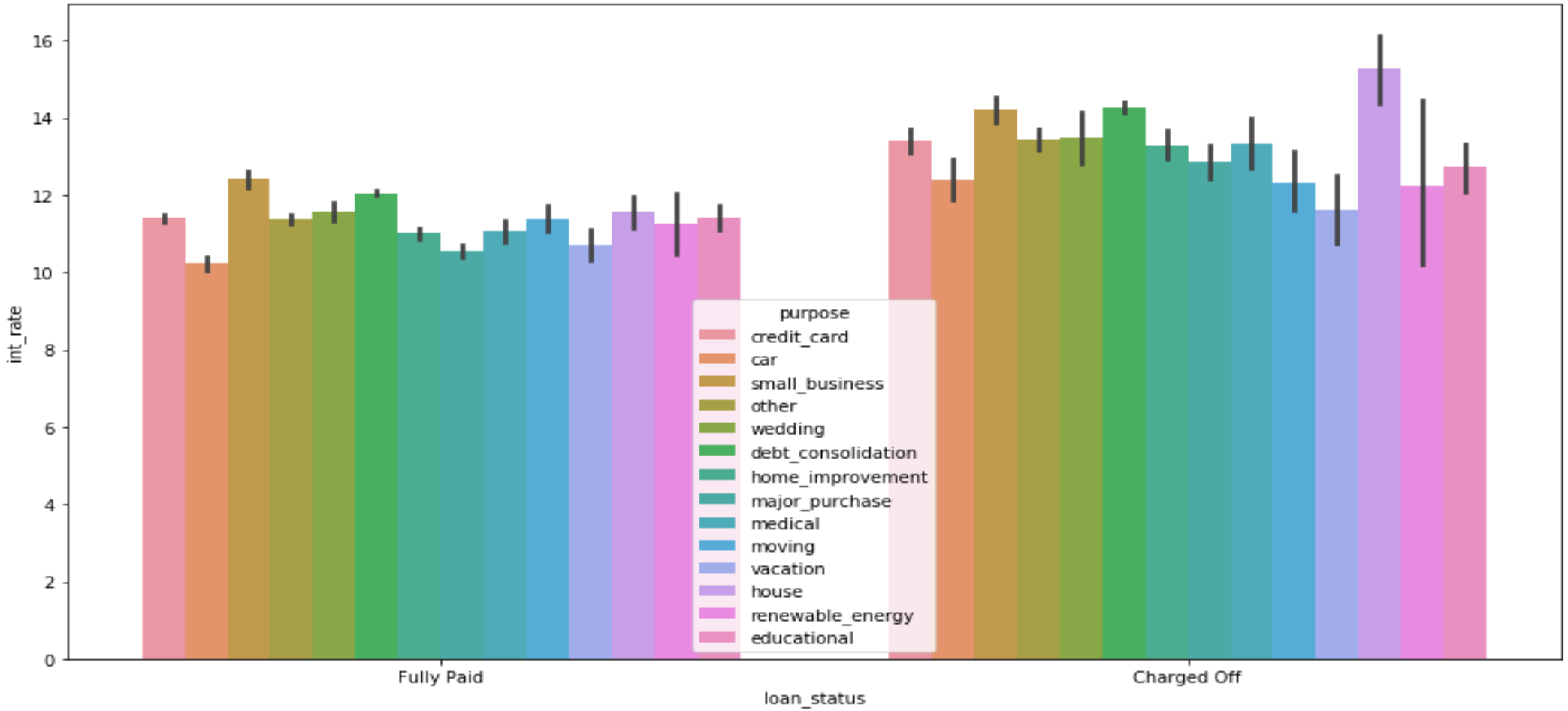
- Another insight which we could come up with after the analysis was that the most number od loans were given in the 10%-15% interest rate bucket **(~48%)**. See first graph below.
- Another insight gathered from the data analysis when we compare the status of the loan Vs the purpose of the loan, we find that the majority of the people **(47%)** take the loan for debt consolidation and out of those **15% default** in their loan payments. See second graph below.



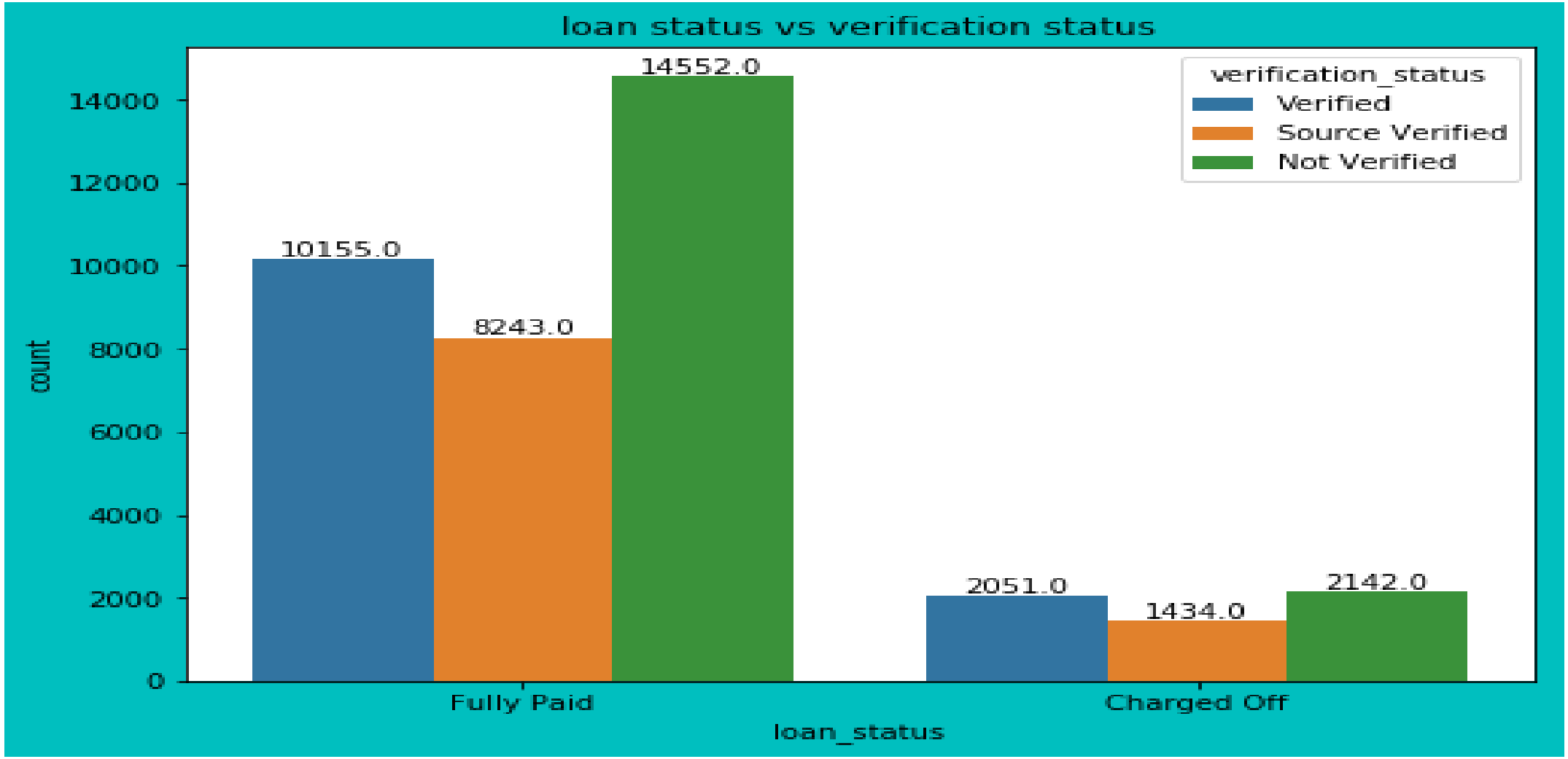
# Solution – contd.

- Further more if we break the dataset even further if we come to know that the House loan, Debt consolidation and Small business are the biggest defaulters. So the company can be more careful in lending out loan for these sectors.



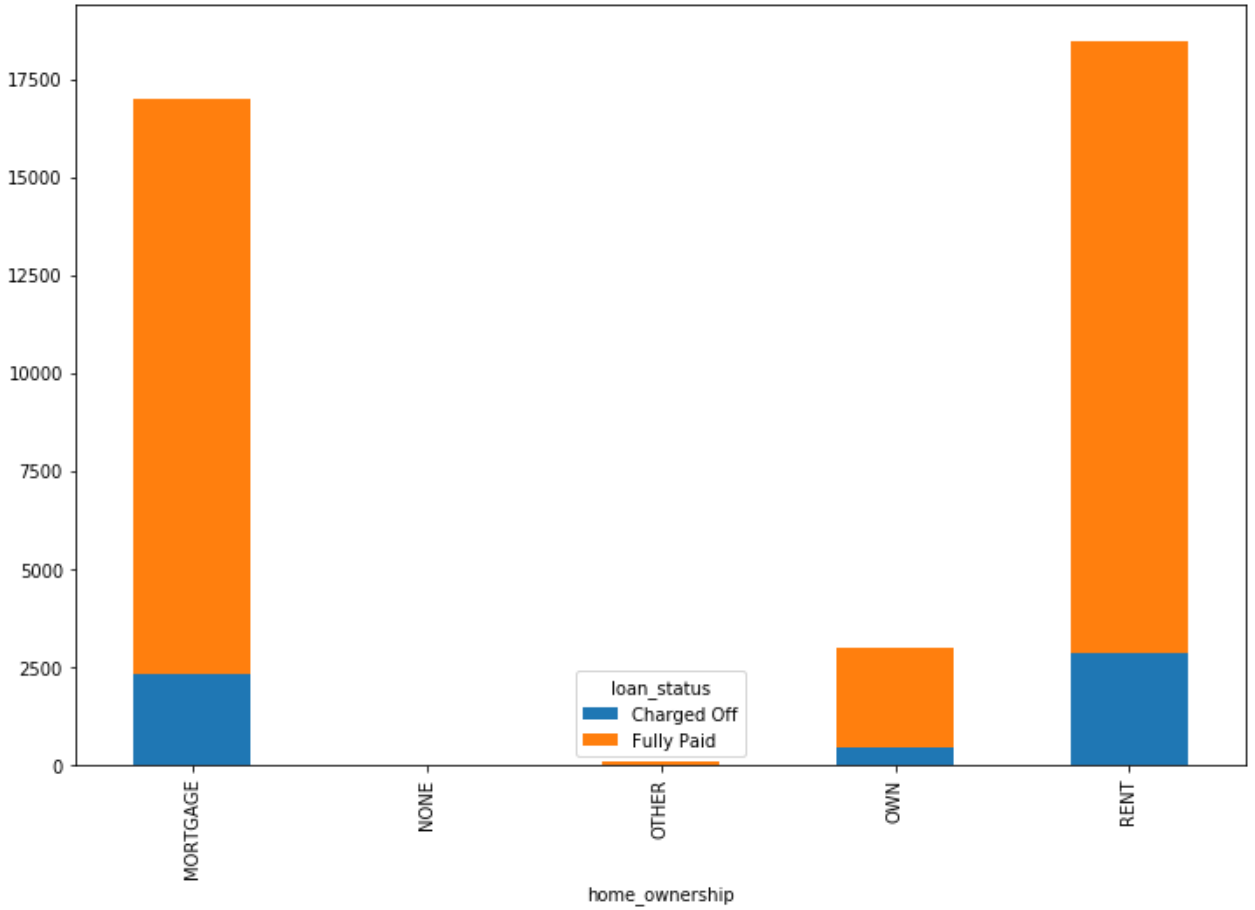
# Solution – contd.

- Another pair of data sets when compared throw the insight that 17% of the verified loans also default which indicates that the company needs to improve their vetting process before issuing the loans.



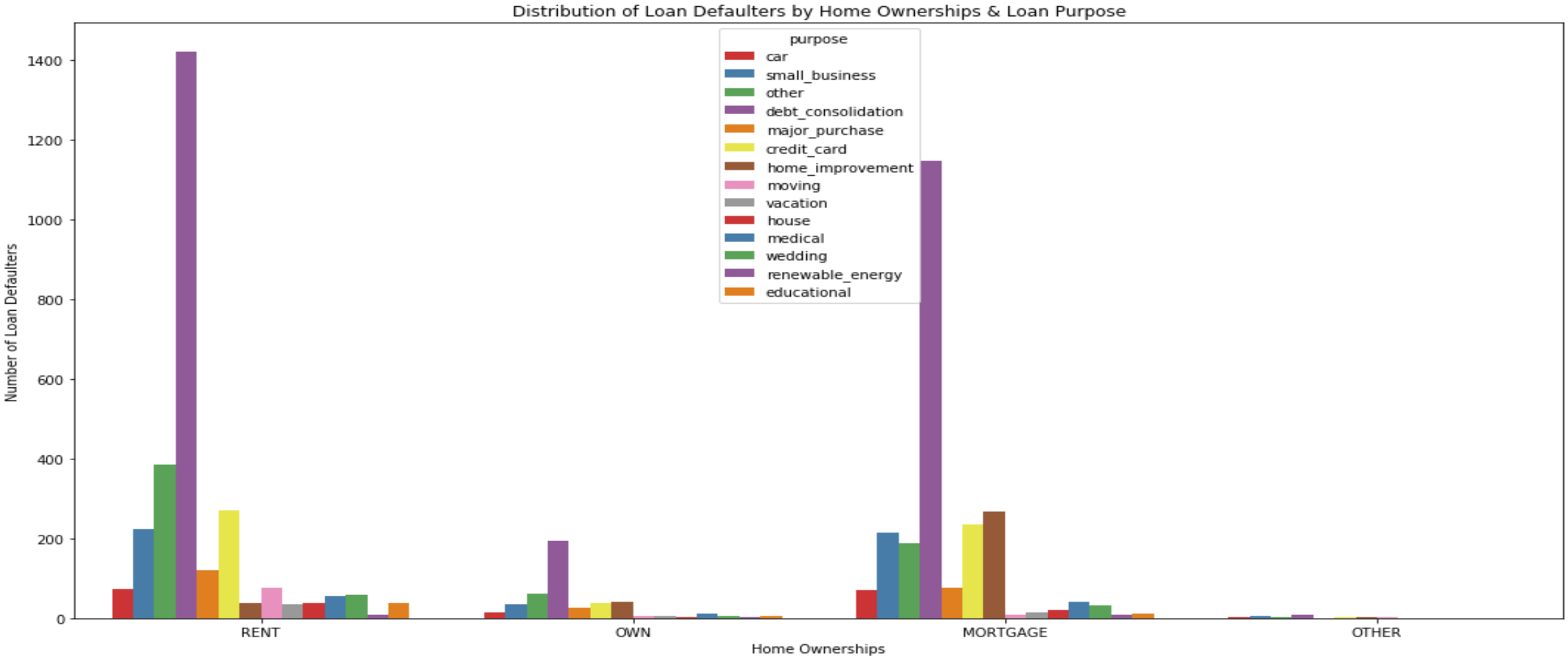
# Solution – contd.

- One more analysis leads us to appoint that the defaulting on the loans increases with increase in the interest rates. This is irrespective of the loan amount bucket. See first image
- Another trend which we could identify was in Home loans category, people who owned the house instead of living in a mortgaged or rented house were less likely to default on their loans. See second image below.

# Solution – contd.

- One last analysis which we carried that we carried out lead us to the insight that
  - for Rented home customers --> highest defaulters are in debt consolidation ,credit card, small business and others
  - for customers living in their own home--> debt consolidation appears to be highest default group
  - for customers in mortgaged homes--> highest defaulters lie in debt consolidation ,home improvement, credit card group



# Solution – contd.

- Finally our analysis helps us to arrive at the following points to help us reduce the loan defaulters:
  - Increase vetting of verified loans as still a significant amount (17%) of the verified loans are defaulting.
  - Try to issue more loans in the lower interest rate bucket as loans falling under the higher interest rates tend to default significantly higher.
  - Be cautious in issuing loans for Home, Debt consolidation and Small business as these categories have the highest percentage of defaulters.
  - For home loans people already living in their own houses are less likely to default but people living in rented or mortgaged house have a higher defaulting percentage. So better checks are to be placed while lending money to people living in mortgaged/rented houses.

# Thank you