EDA Case Study

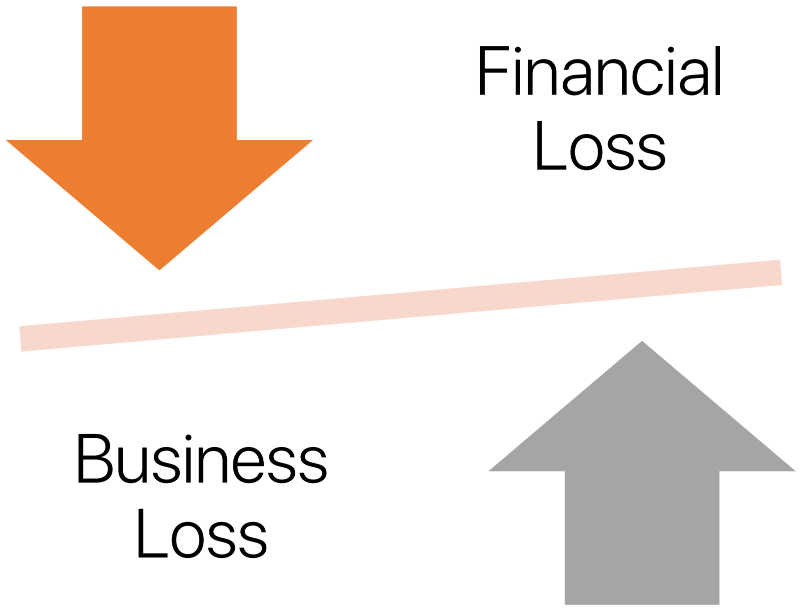
PG Diploma in Machine Learning and AI March 2018

Bishwarup Bhattacharjee (F)
Akansha Jawanpuria
Ravi Kumar N K
Sabyasachi Dey

Abstract

In this presentation, we are going to analyse the lending behaviour of a consumer finance company that specializes in lending various types of loans to urban customers.

Problem



One of the major problem in front of a consumer finance lending company is to filter out the bad loans i.e. the loans that are risky and likely to get charged-off causing the company a financial loss. However, if the norms for made too stringent, that might result in potential business loss i.e. not approving loans to eligible customers. A balance has to be struck somewhere in the middle.

Solution

We start with our domain knowledge to form certain hypotheses about the potential indicators of loan defaults

We validate our hypothesis by analyzing the dataset and making representative graphs and charts

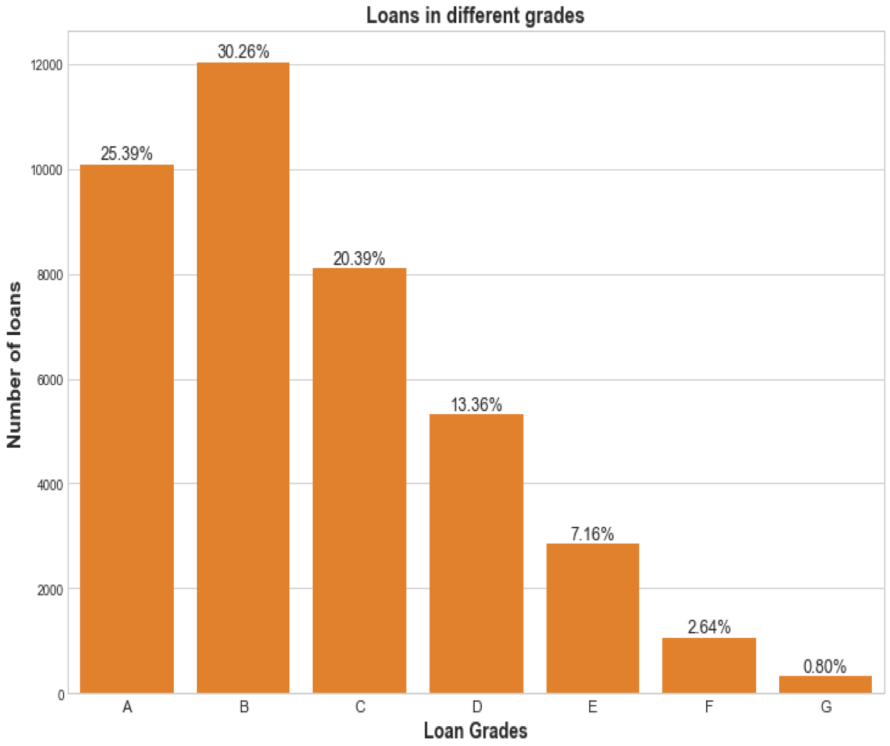
From the data we figure out the drivers of loan defaults and advise the business to act accordingly to minimize the risk of default.

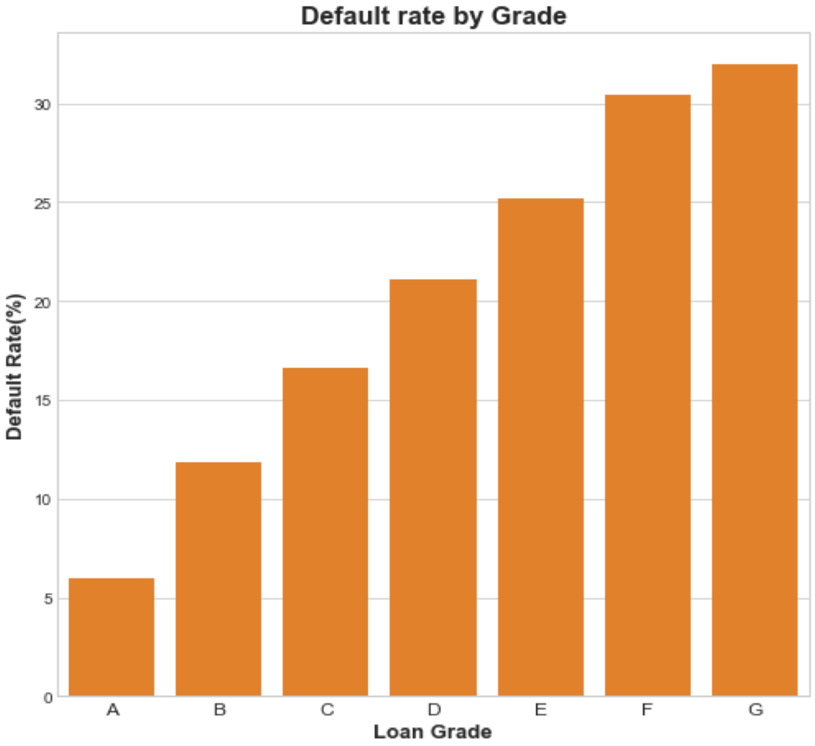
The Dataset

39,717	**2007 - 11**	**39**	**14.17%**
Applicants	Time period	States	Historical Default Rate

Our Findings

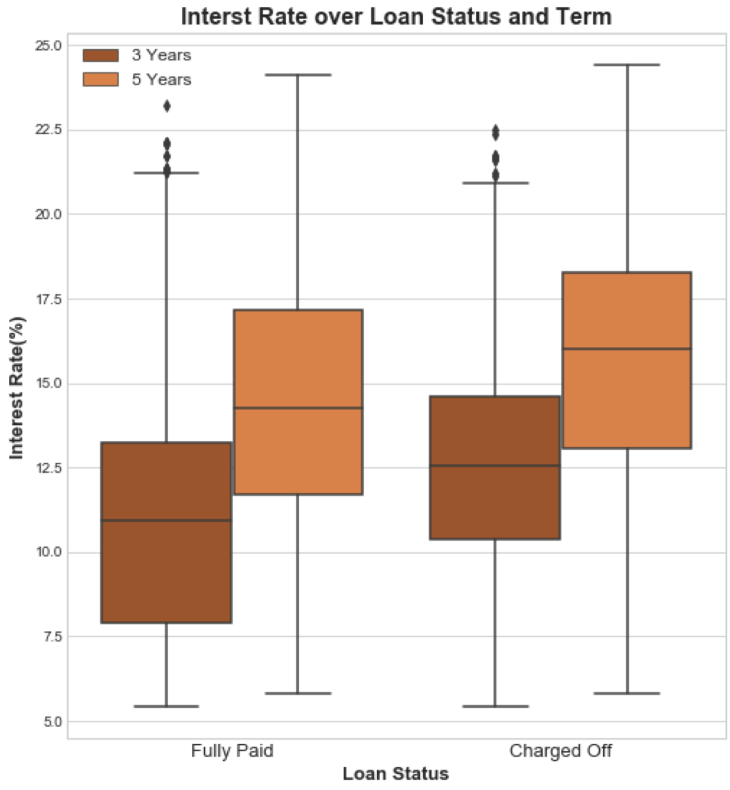
Loan Grade

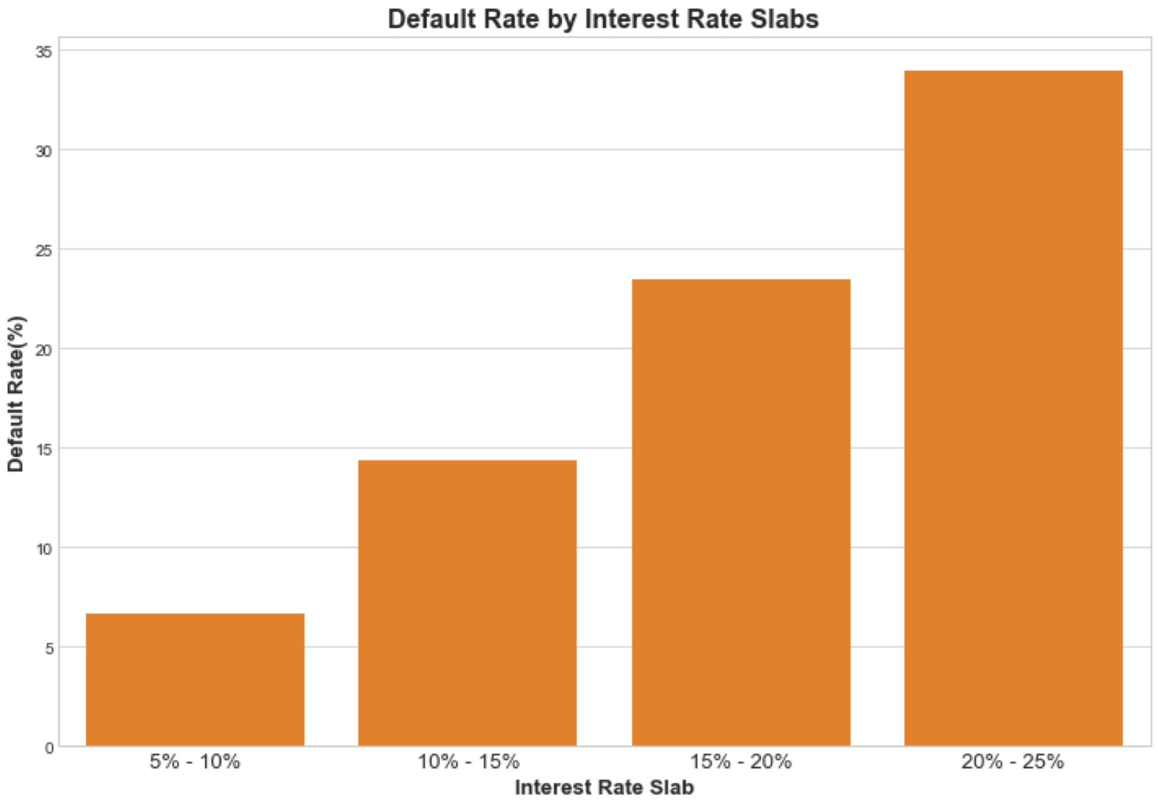




Loan grade is a grade assigned by the company to the loans it issues that denote the risk factor attached with it. In our data we have 7 loan grades from A through G – A being the least risky and G being the most. As we see from the graphs above though the number of loans in the risky groups are much less – they also have a very high default rate. From D onward the default rate is over 20% which goes well over for grades F and G.

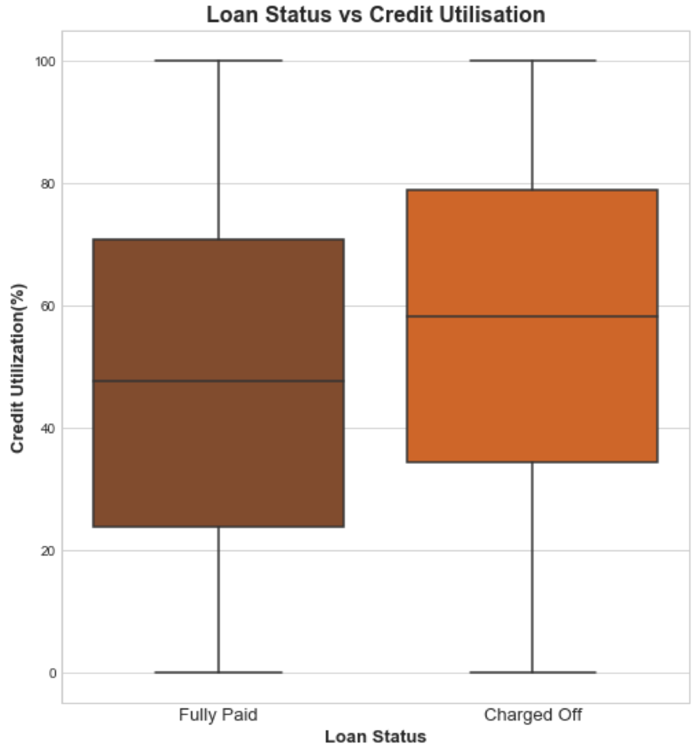
Interest Rate

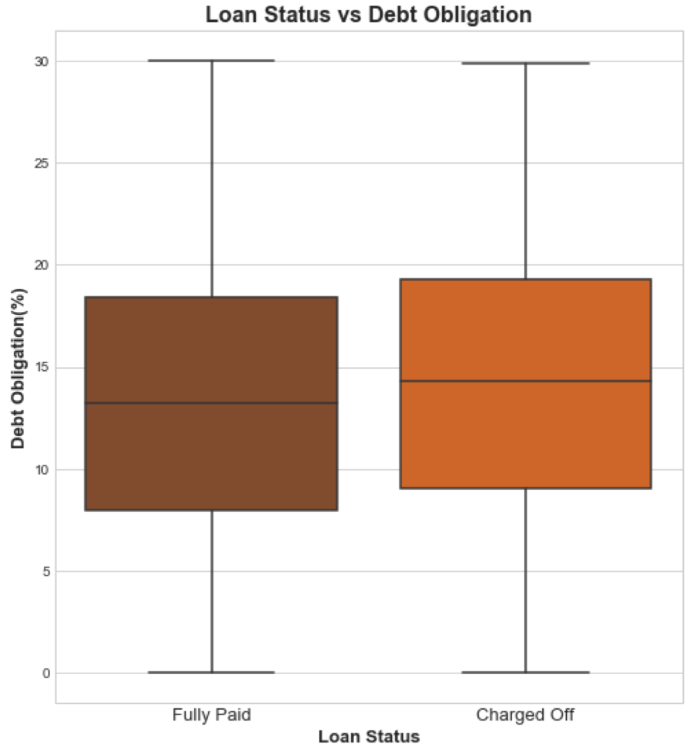




We also find out that interest rate has potential impact on default rate. For both repayment tenures of 3 years and 5 years, the interest rate levied to charged-off loans were substantially higher than that to paid-off loans. We also observe how fast and steadily the default rate goes up with increase in interest rate. While the default rate in the 5-10% interest rate bracket is as small as ~5%, the same goes above 30% in the 20-25% interest rate bracket.

Credit Utilization and Debt Payable

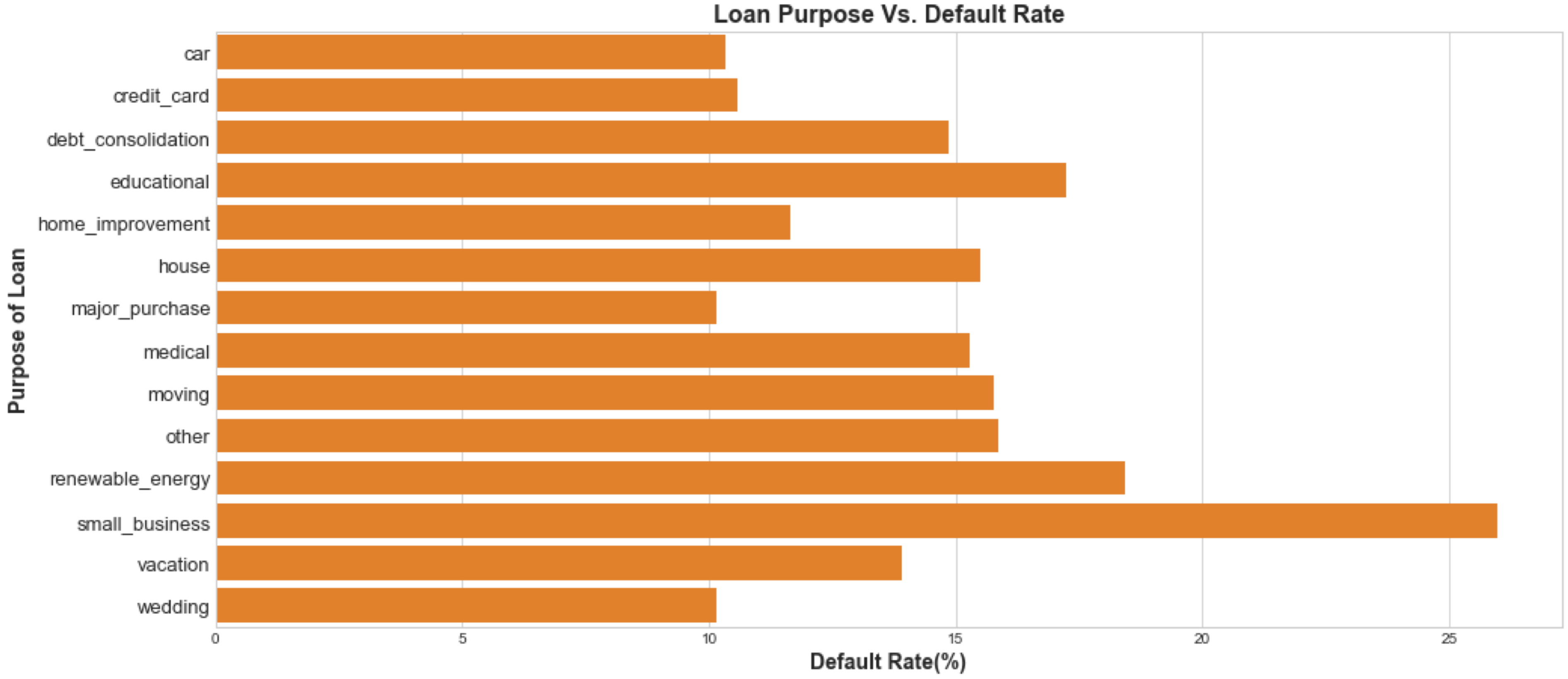




Credit Utilization is the proportion of credit that the borrower has historically been utilising out of total credit available. A high credit utilization points to financial instability due to which we can see the charged off loans have a higher credit utilization compared to that of fully paid loans.

Similarly, the debt to income ratio measures the part of the income that the applicant has to spend towards their existing debt obligations. Quite intuitively, a higher ratio means heavier debt obligation and less financial flexibility. Hence, just in line with our previous observation, we see the borrowers who have defaulted in the past had a higher debt to income ratio on an average.

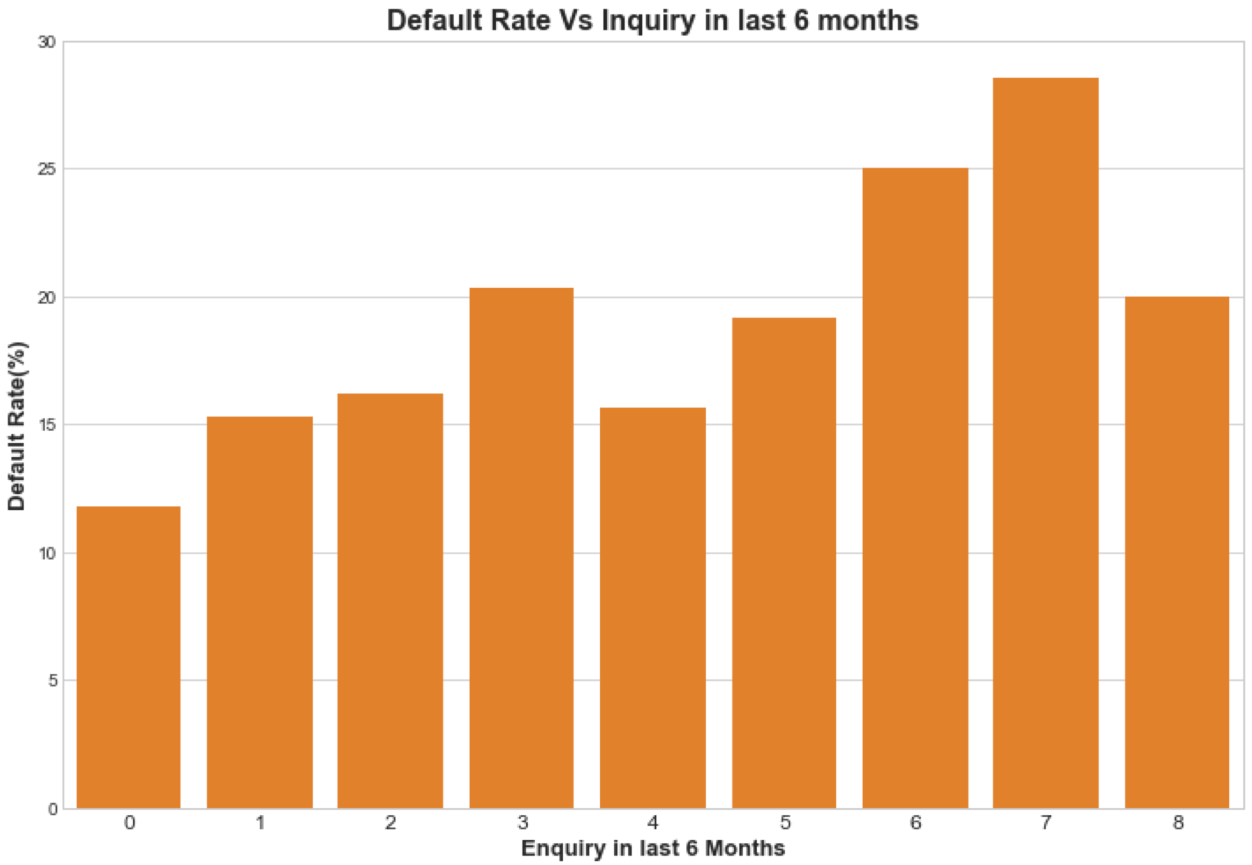
Loan Purpose



We discover the fact that loan purpose has a significant impact on the probability of default. From the above graph we see that loans issued for specific reasons carry differential risk. Loans issued for constructing small businesses are the most risky in that they have got over 25% default rate. One plausible reason for this incident may be due to the fact that business owners plan to pay off the loan with their return from the business, but in practice many of such businesses fail in the future rendering the borrower incapable of paying their debt. So it’s really important for LC to assess the risk factor of such businesses prior to disbursement of the loan.

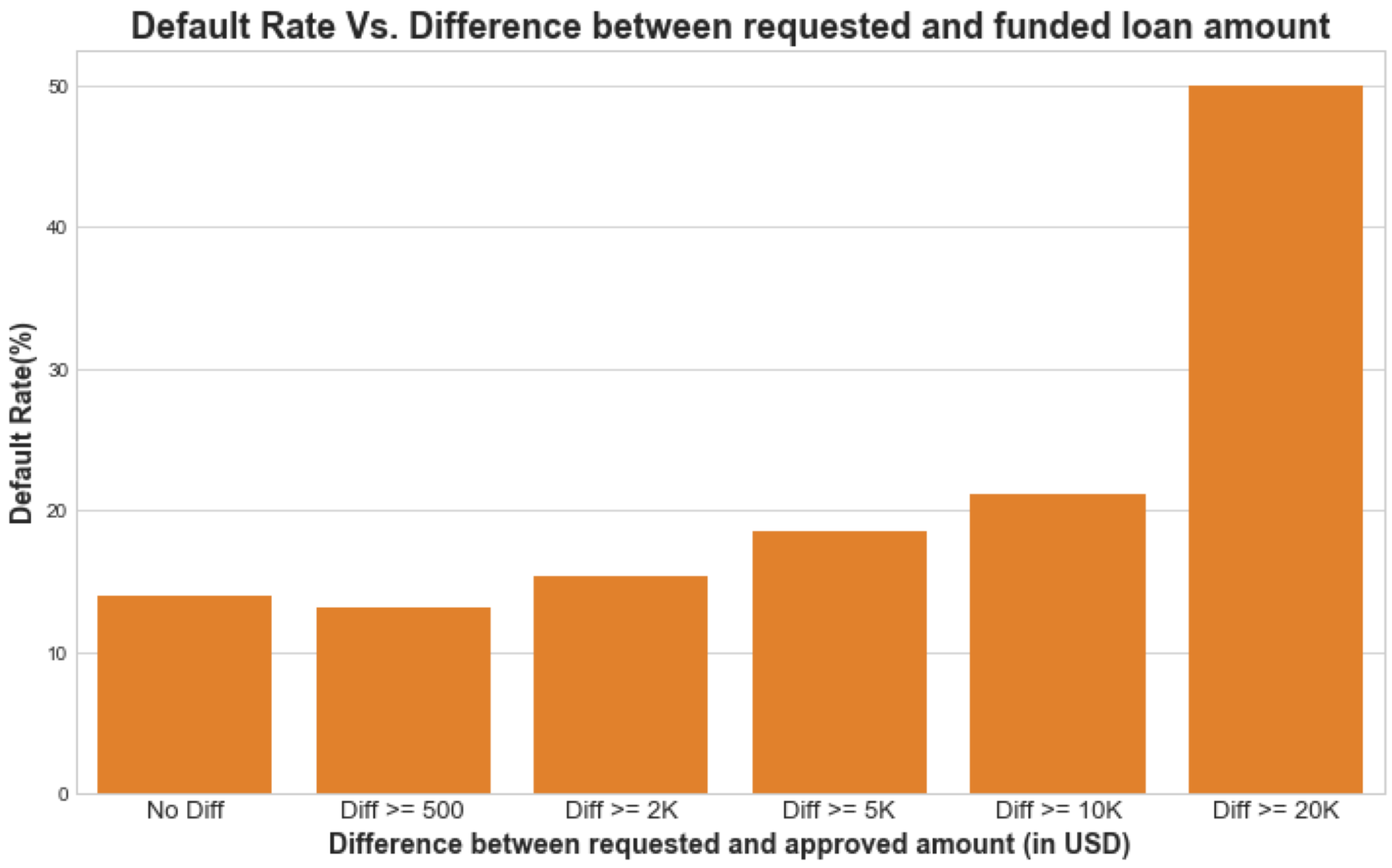
Educational and renewable energy loans have also got unusually high default rates.

Credit Inquiry in last 6 months



The number of credit inquiries reflects on the credit hungriness of the borrower which in turn draws upon the financial instability that the borrower has been experiencing lately. A higher count of inquiry in the last 6 months while applying for a loan can also potentially mean that a number of financial institutions have already rejected the loan application recently. These applicants are admittedly more vulnerable to default. Our finding is also in line with that. In the graph above we see how the rate of default keeps increasing overall with growing number of inquiry.

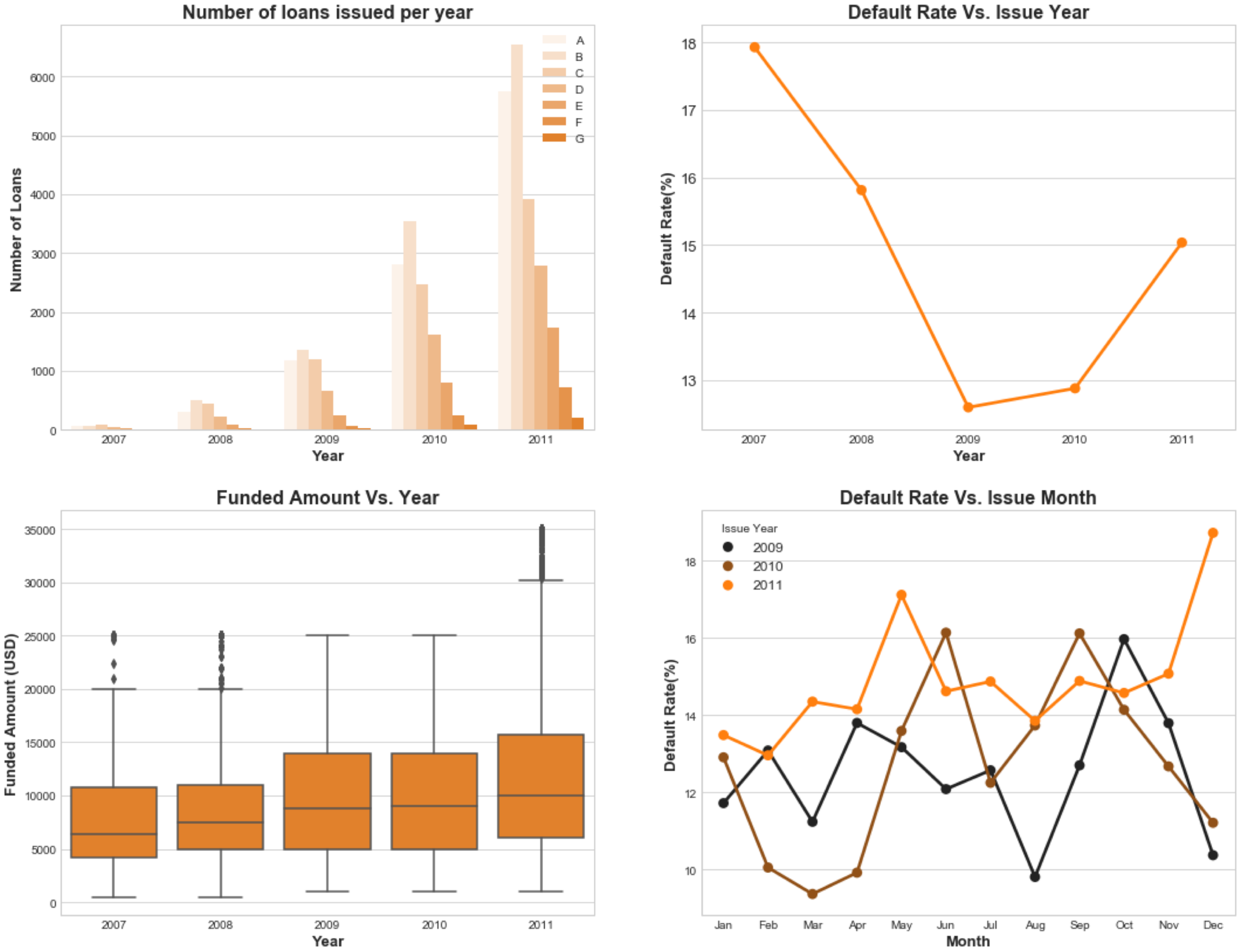
Requested Amount Vs Approved Amount



This is often the case when the borrower requests for a specific amount but LC only approves a certain fraction of that amount. We analysed the borrowers for whom the LC approved a lesser amount and the result was quite interesting.

In the graph, we can see how the rate of default increases with the difference of requested amount and approved amount. In the extreme case where the approved amount is less by USD $20K or more, the default rate skyrockets to 50%.

A lookback of how the LC evolved over time...



With the tiny amount of loan disbursed in 2007, we can assume that that is when the company came into operation. To start with the company did pretty well to fight off against the *bad* loans but in 2011 the default rate took off again.

This has to do with the greedy lending that LC had gone through at that time period. The long upward tail of the boxplot (bottom-left) corresponding to the year 2011 provides evidence to our claim.

In 2011 December, the default rate was peaked at 3 years' high. We analysed and figured out that the average lending amount during that month was much higher than the year's average.

Advice

- LC should not go all the way out of its usual lending pattern and should not approve very high loan amounts without proper scrutiny.
- The company should try to minimize the loans even further in the risky grades e.g. E, F and G.
- They should pay special attention while approving loans for purposes like small business, renewable energy and education.
- The company should be extra cautious when approving an amount less than the requested amount.