EDA Case Study

Group Members:
Shwetank Pandey
Vijaya Gupta
Nirmal Maheshwari
Vijay Mane

Abstract

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. Lending loans to ‘risky’ applicants is the largest source of financial loss (called credit loss)

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.

Company wants to understand the **driving factors (or driver variables)** behind loan default, i.e. the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment

Problem Solving Methodology-Analysis Approach

1. Sourcing the data.
2. Cleaning the data and outlier removal of the annual income column.
3. Segmented Univariate/ Univariate Analysis
4. Bivariate Analysis
5. Finding the relations between important columns

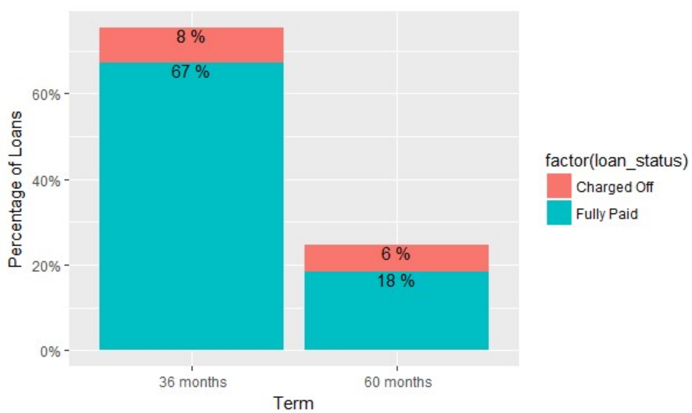
Funded Amount Analysis

1. From the summarization we will get to know that average funded amount of the Charged off loans are more then the average funded amount of the Fully Paid Loans. This shows that as the amount increases it tends to default more as the EMI for that loan also increase.
2. Deriving a new column subtracting the amount given by investor and the total amount given will give you the amount given by the company itself, and it shows from the summarization if the avg amount invested by the company for charged off loans is more then the fully paid loans.

Status of Loan	Average Funded Amount	Average Funded Amount by Company
Charged Off	11676	883
Fully Paid	10505	500

Interest Rate, Term and Amount Relation

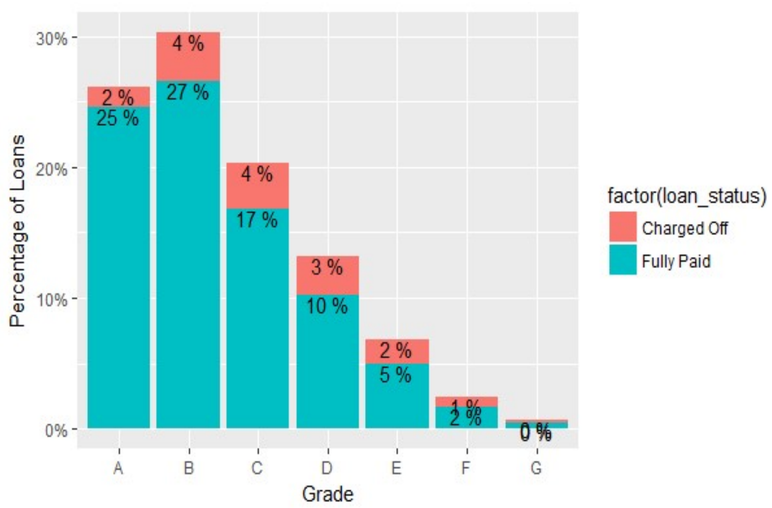
1. Below graph will show that there are 36 month term loans tend to default more than 60 month term loans.
2. The summary metrics below will show that the Average interest rate and Avg. funded amount is more for the 60 months loans still they tend to default less then 36 month showing that the term of loan is an important factor, even if the interest rate is less the small term loans tend to default more.
3. Though the average interest rate for the charged off loans comes out to be 2 % more then the Fully Paid Loans from the below summary metrics.



Term	Loan Status	Avg Funded Amount	Avg. Interest Rate
36	Charged Off	9176	12.36
36	Fully Paid	9394	10.81
60	Charged Off	15038	15.73
60	Fully Paid	14582	14.41

Grade, Interest rate, Funded Amount Relation

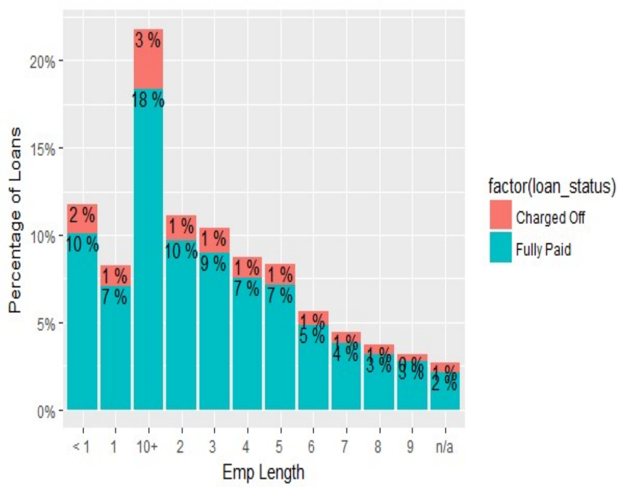
1. From the blow graph, the grade B and C loans tend to default more than the other grade Loan.
2. From the summary metric below, it shows that the Average interest rate and Average funded amount increases along with the grade, but the B grade loans provides a little bit higher amount at a less interest rate and most of the people gets this grade loan, and the C grade loan offers the same amount as B grade but a higher interest rate, so people getting a C grade loan tend to default more as the interest rate is higher, and as the more people gets B grade loan because of higher loan amount at lower interest rate, they tend to default more.



Grade	Average Amount Funded	Average Interest rate
A	8343	7.32
B	10594	11.00
C	10500	13.54
D	11821	15.67
E	14942	17.61
F	17200	19.63
G	19618	21.28

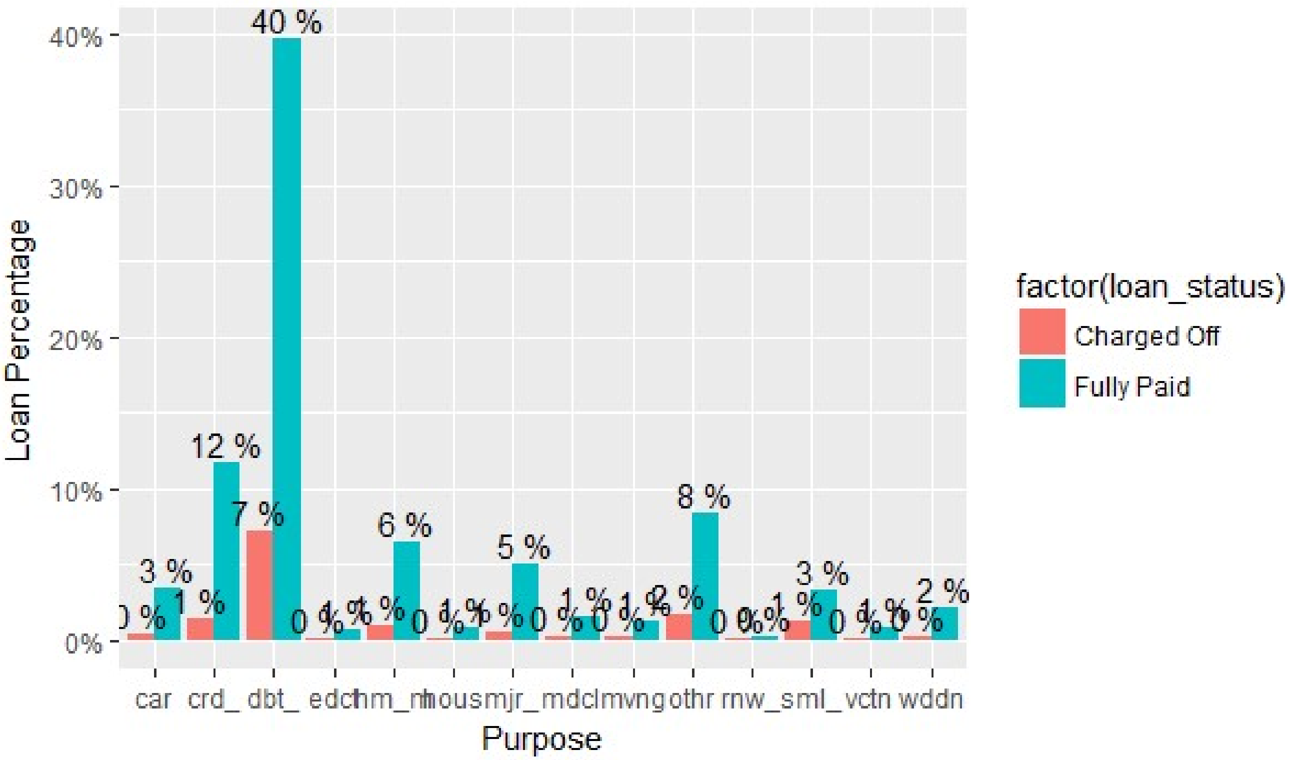
Employment Length And Annual Income

1. Below graphs shows that the customers with employment length more than 10 years tend to default more than others.
2. The summary metrics will show that the income is higher for the persons with 10 plus years experience (74945) but accordingly they might be having more liabilities then other employment groups hence tend to default.
3. The average annual income of the charged of loan comes out to be less than the average annual income of the fully paid loans.



Loan Status	Average income
Charged Off	59465
Fully Paid	65674

Purpose and Income and DTI

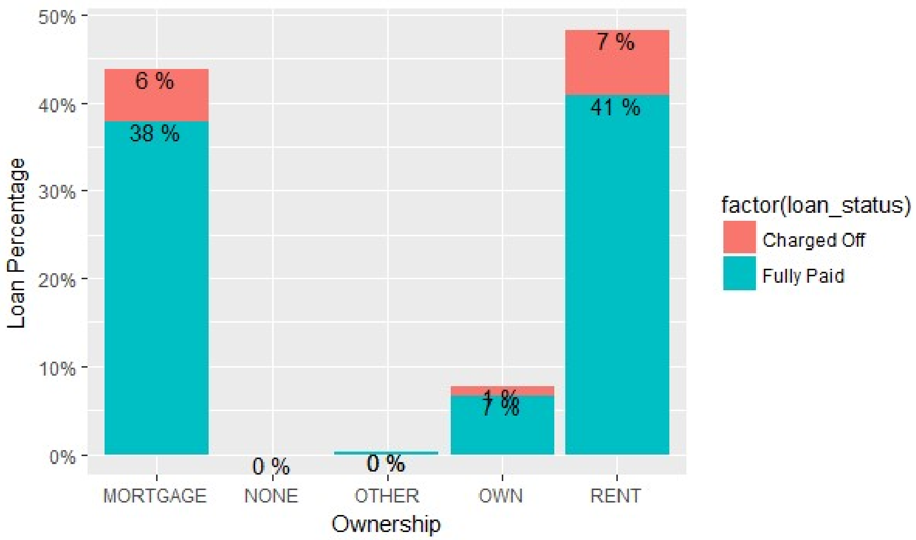


1. The graph shows that the persons who has taken the loan for **Debt consolidation** tend to **default much** more then the others.
2. The metrics will show that the loan taken for **Debt Consolidation has higher interest rate** than others hence tend to default more and the Debt to income ratio is much higher for these loan then others

Purpose	Avg Income	Rate	Dti
Car	60708	10.45	0.111
Cr Card	67425	11.60	0.146
Dbt Cons	63990	12.38	0.144
Edu.	51760	11.65	0.11

Home Ownership and Annual Income And DTI

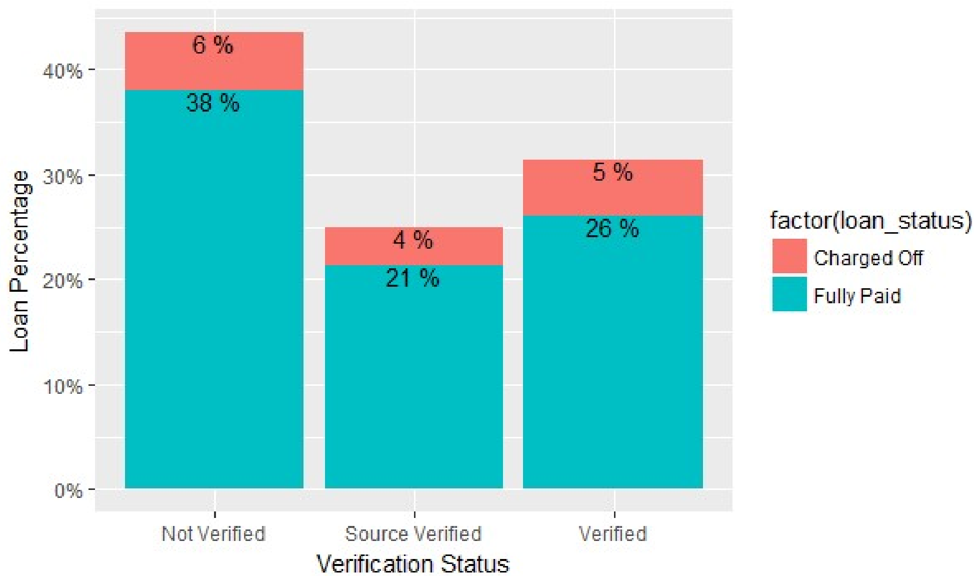
1. From the graph it is shown that the loan taken by customers having rented house or a mortgage tend to default more then the customers having their own house.
2. Summary metrics will show that the customers having rented house will be having comparatively less average annual income and more debt to income ratio, the DTI for mortgage homes is also more.
3. This is because the loan taken by people having rented and mortgage houses is for debt consolidation which tend to default.



Home Ownership	Avg. income	Debt To Income Ration
Mortgage	76793	0.132
Other	64003	0.11
Rent	55513	0.13
Own	55350	0.13

Verification Status

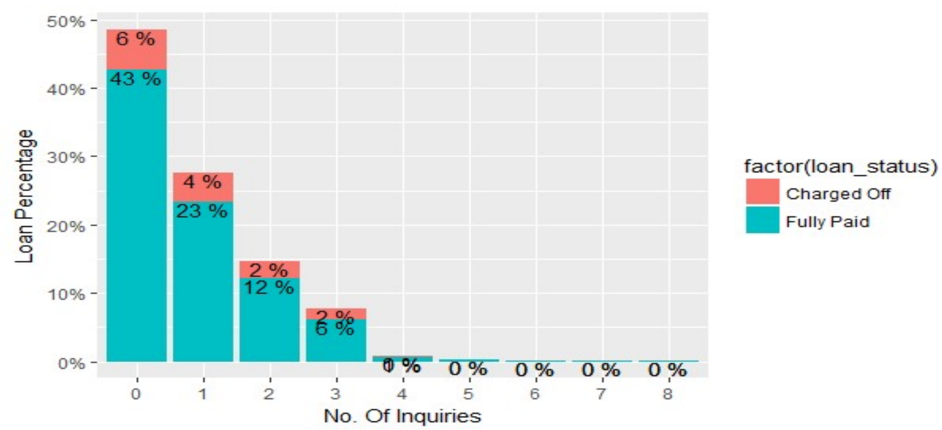
1. Graph Shows that the loans for which the income source is Not Verified tend to default 2 % more then if the Source is verified.



Revol Balance And Revol Util And Inquiry In Last 6 months

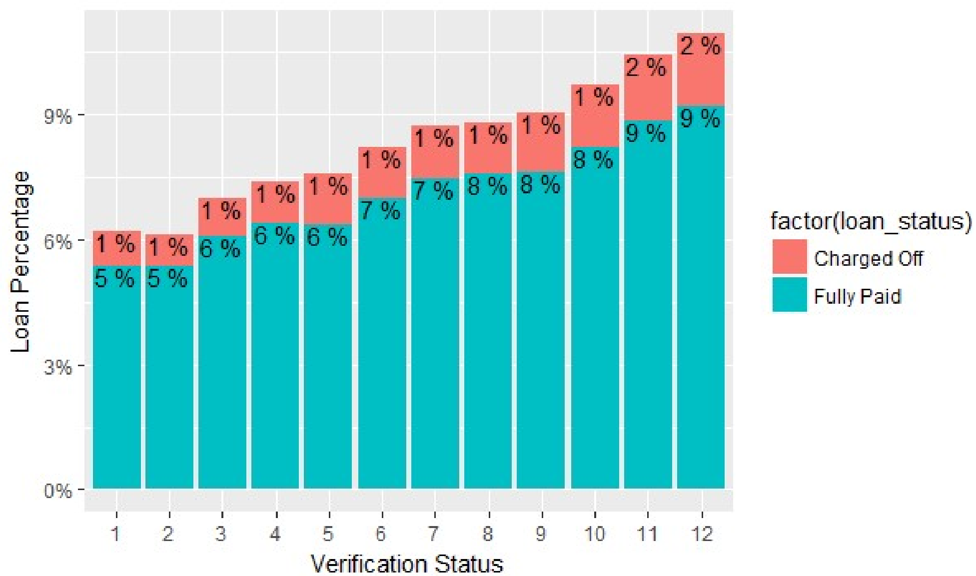
- Deriving a new column from the given columns gives the amount of credit the borrower is using relative to all available revolving credit.
- The summary metrics will show revolving amount utilized by the customer who defaulted is more then the ones who pays the loan.
- The graph shows that the loans for which the inquiries are none tend to default more then for which the inquiries are made.

Loan Status	Average Amount
Charged Off	8632
Fully Paid	7485



Issue Date

- The graph shows that the loans issued in the month of November and December tend to default more then the others, the reason can be as these months are festive months and the loans taken in these months tend to have higher interest rate comparatively.



Conclusions/Driving Factors

Main driving factors from the above Analysis

1. More the Funded Amount and Funded amount by investors, more it tends to default.
2. More the interest rate & less tenure more the default rate, especially grade B and C loans tend to default more.
3. Purpose is a driving factor, Debt Consolidation loan tends to default more due to high Debt To Income Ratio.
4. Home Ownership is a driving factor as well; because the people with Rented homes have less annual Income comparatively.
5. More the utilization of the revolving credit more the customer tend to default.
6. If the source of income of the customer is not Verified the default chances are 2 percent more.
7. If the Employment length grows the customer tend to default more may be because of the grown responsibilities. Also, when the employment length is below a year it does tend to default.
8. Issuance Month for loans can also be a driving factor.