

Leading Club Case Study

Group Members:

Vishnu CV

Nivetha

Sl no	Content	Slide No
1	Understanding Business Problem and Requirement	3
2	Data Cleaning.....	4
3	Data Analysis.....	
	3.1. Univariate Analysis.....	5
	3.1.1 Analysing loan Acceptance scenario.....	6
	3.1.2 Univariate plots.....	7
	3.2. Bivariate Analysis.....	15
	3.2.1 Bivariate Plots.....	16
	3.2.2 Identify the key driving factors.....	20
4	Conclusion.....	21

1. Business Problem and Requirement

You work for a consumer finance company which specialises in lending various types of loans to urban customers

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.

When a person applies for a loan, there are two types of decisions that could be taken by the company:

Loan accepted: If the company approves the loan, there are 3 possible scenarios described below:

1. Fully paid: Applicant has fully paid the loan (the principal and the interest rate)
2. Current: Applicant is in the process of paying the instalments, i.e. the tenure of the loan is not yet completed. These candidates are not labelled as 'defaulted'.
3. Charged-off: Applicant has not paid the instalments in due time for a long period of time, i.e. he/she has defaulted on the loan

Loan rejected: The company had rejected the loan (because the candidate does not meet their requirements etc.). Since the loan was rejected, there is no transactional history of those applicants with the company and so this data is not available with the company (and thus in this dataset)

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. In other words, the 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 utilise this knowledge for its portfolio and risk assessment

2. Data Cleaning

- The data provided were having significant number of null values for each column. So as a first step, total null value percentage of each attribute was calculated and then removed the attributes which were having more than 50% null values.
- Column which were having same values were identified and removed.
- Few columns which were non relevant for the analysis were removed.
[Id, member_id, Url, desc, Title, Zip_code, Addr_state]
- Also removed columns which are part of post loan approval.
[last_pymnt_amnt, last_pymnt_d, last_credit_pull_d, collection_recovery_fee, recoveries, total_rec_int, Total_rec_late_fee, total_rec_prncp, total_pymnt_inv, total_pymnt, revol_bal, Out_prncp]

After cleaning data set of size 36502 x 26 was obtained.

Univariate analysis

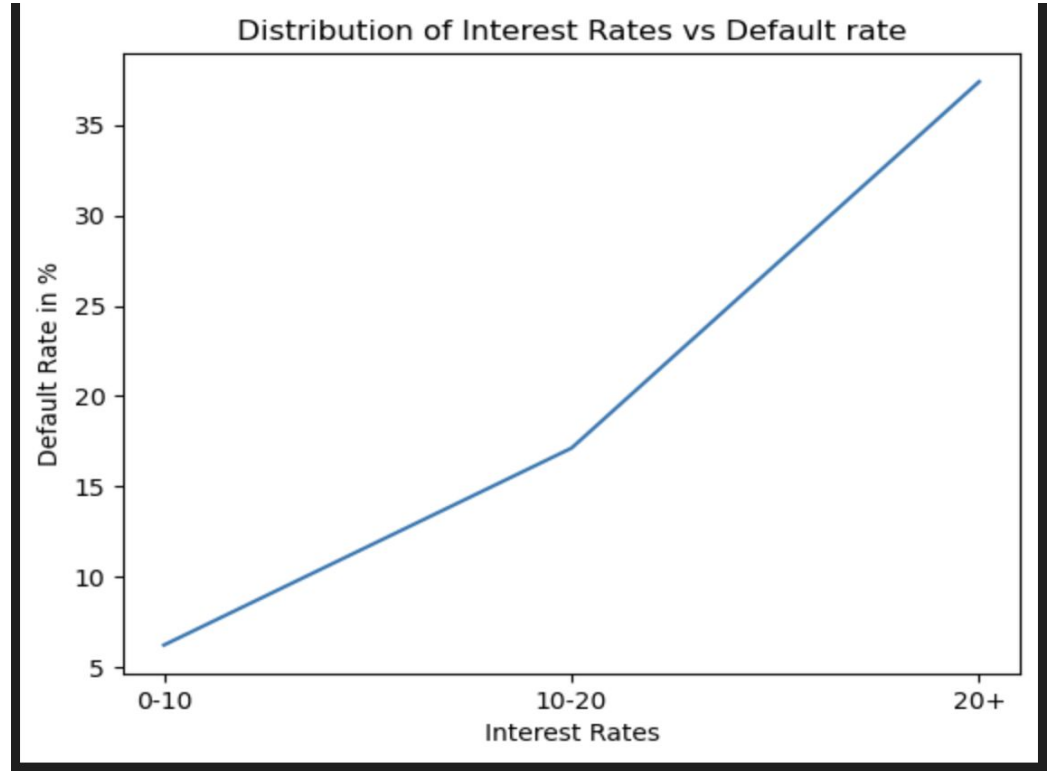
Analysing Loan Accepted Scenarios

- Loan status is the main parameter than can be used to get the key driving factors behind the loan default. We have fully paid, current and charged off scenarios. Since Current is the applicants who are in the process of paying instalments so current can be removed from the data.
- Also, we need to convert the fully paid and charged off scenarios to a quantitative value for analysing the percentages.
- Target is to find the 'default user' so convert loan status to quantitative data
- Assigned 1 for 'Charged Off and 0 for 'Fullypaid'

Distribution of Interest Rate with Default Rate

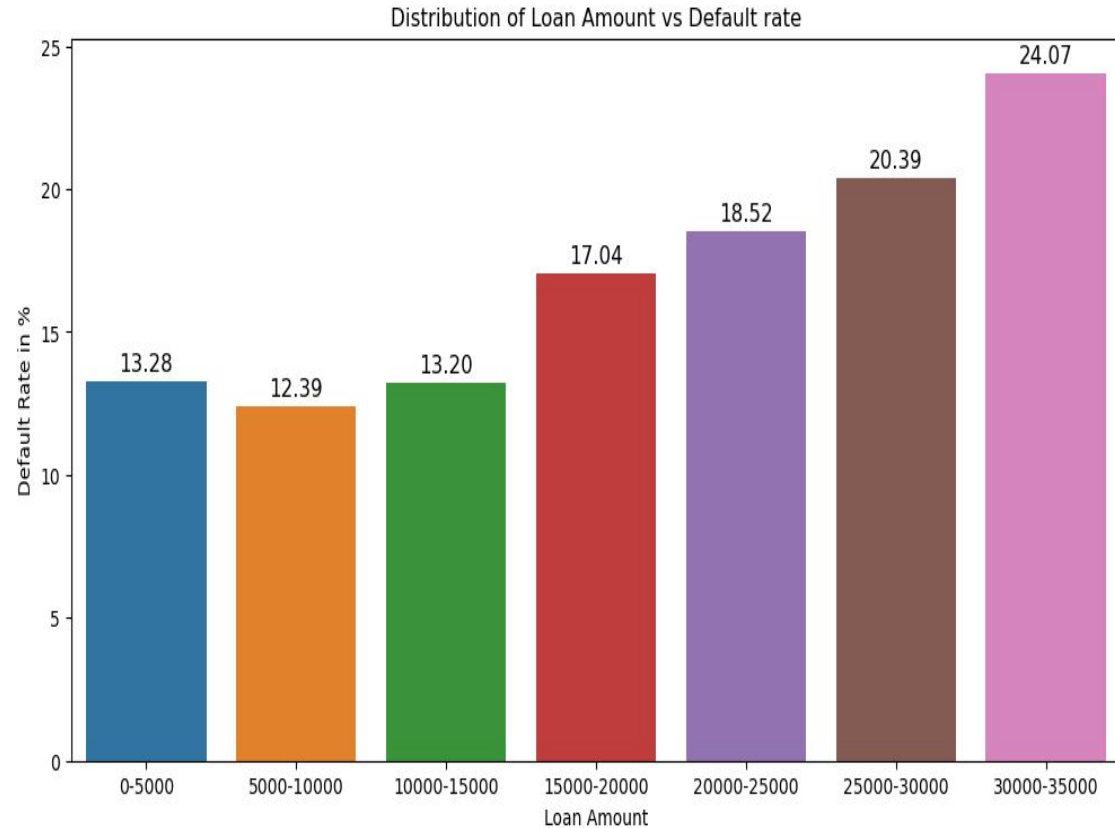
Loans having interest rate more than 20% are likely to be defaulted.

There is a constant increase in the default rate with the interest rate



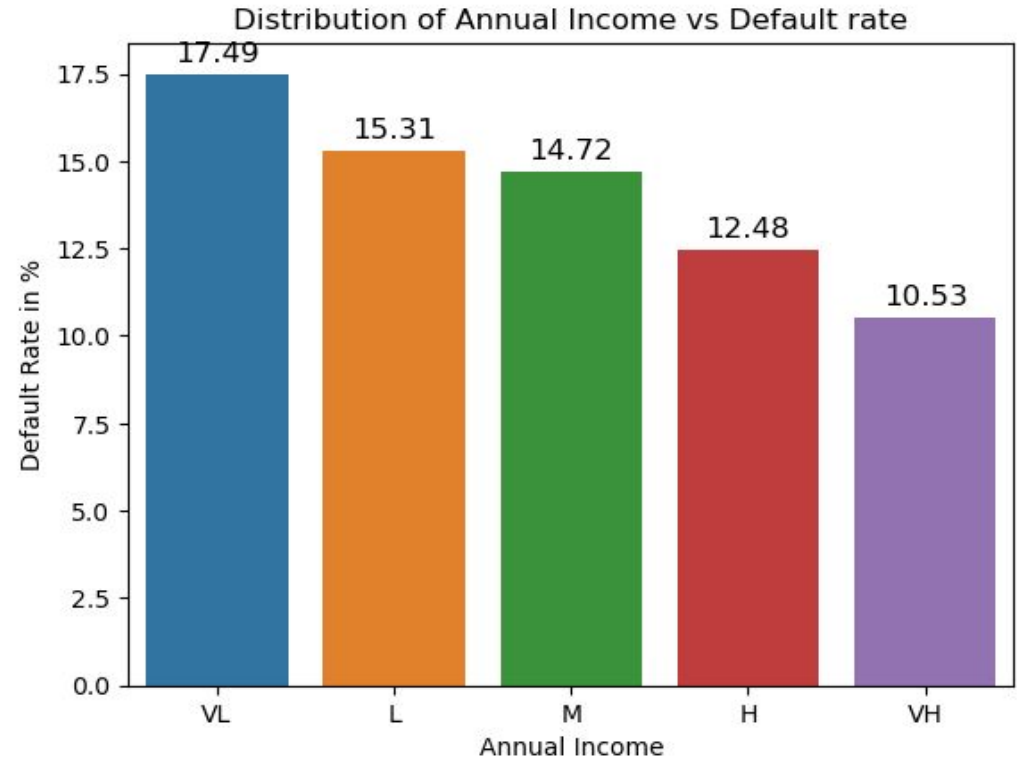
Distribution of Loan Amount vs Default Rate

Loan amounts which are greater than 15k is likely to be defaulted as we can see that there is an increase in the percentage for amount greater than 15k and the amounts less than 15k is likely to have lesser default rate.



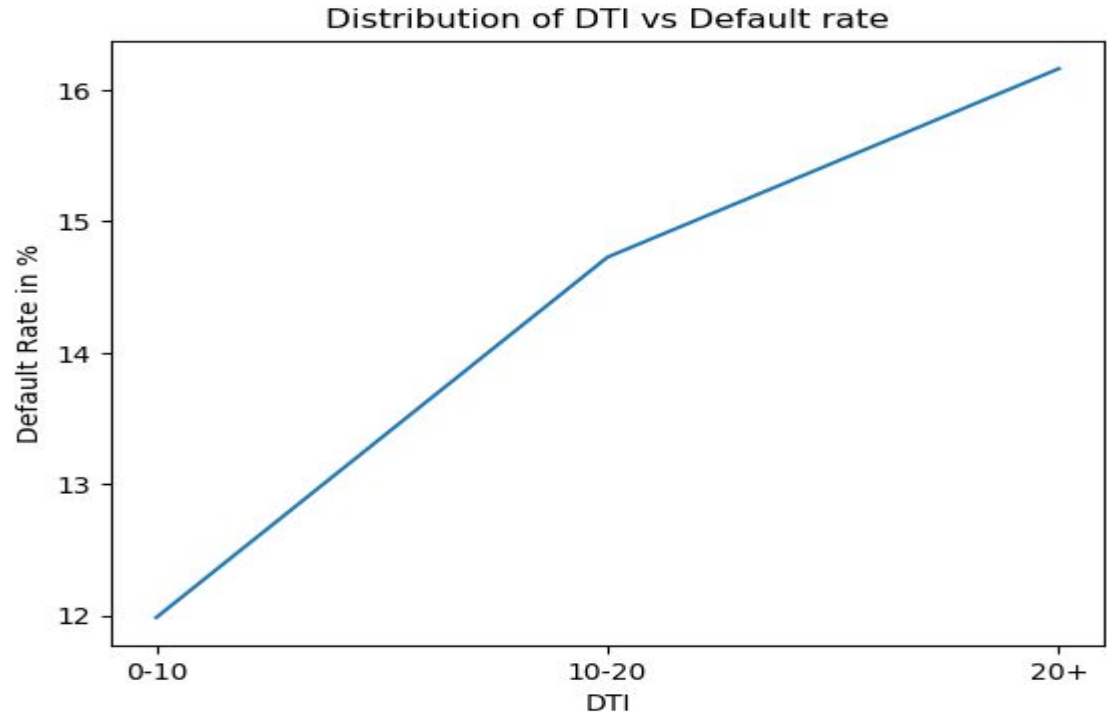
Distribution of Annual Income vs Default rate

Default rate is higher for annual income which comes under very low category. The changes of defaulting is high for people having very low income.



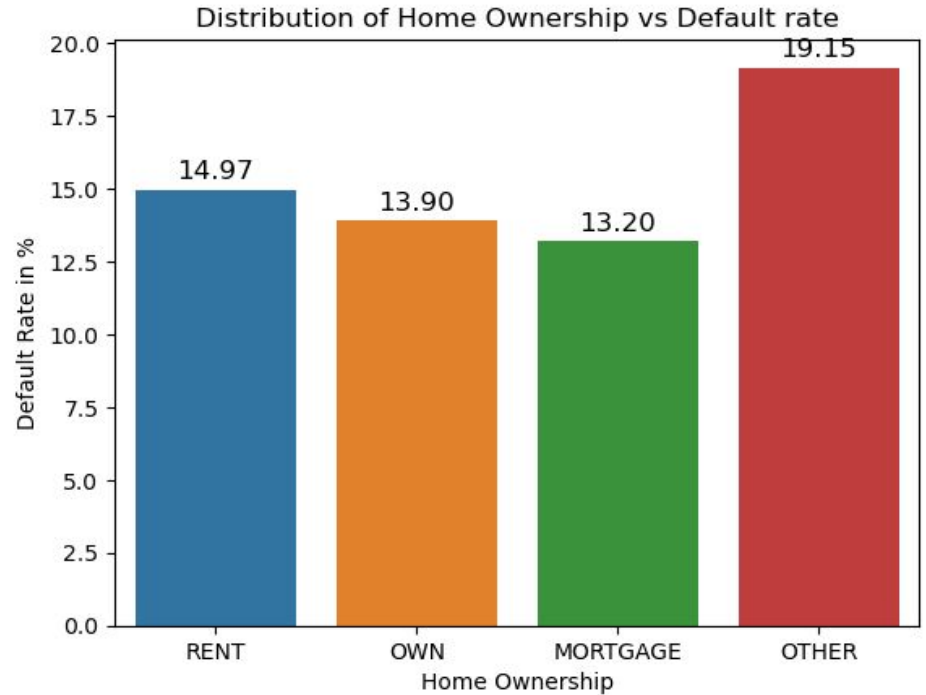
Distribution of DTI vs Default Rate

Default rating is high for dti rate more than 20% which means there is a clear chance of defaulting for higher dti rating.



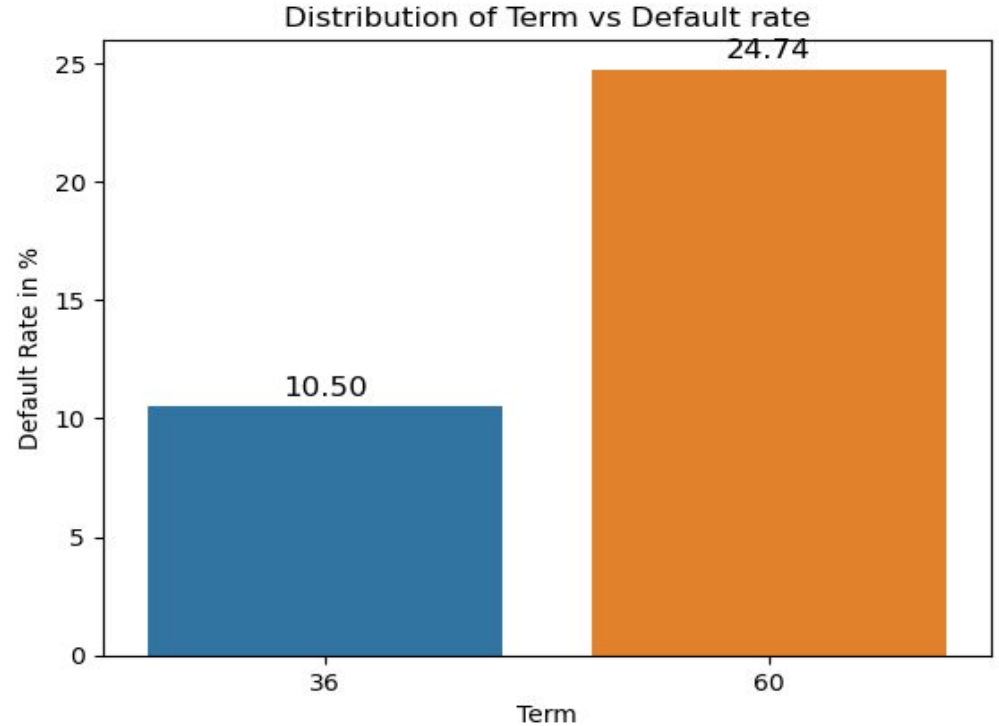
Distribution of Home Ownership vs Default rate

Other category ownership people have higher default rate which means there is a high change of defaulting for other category home owners



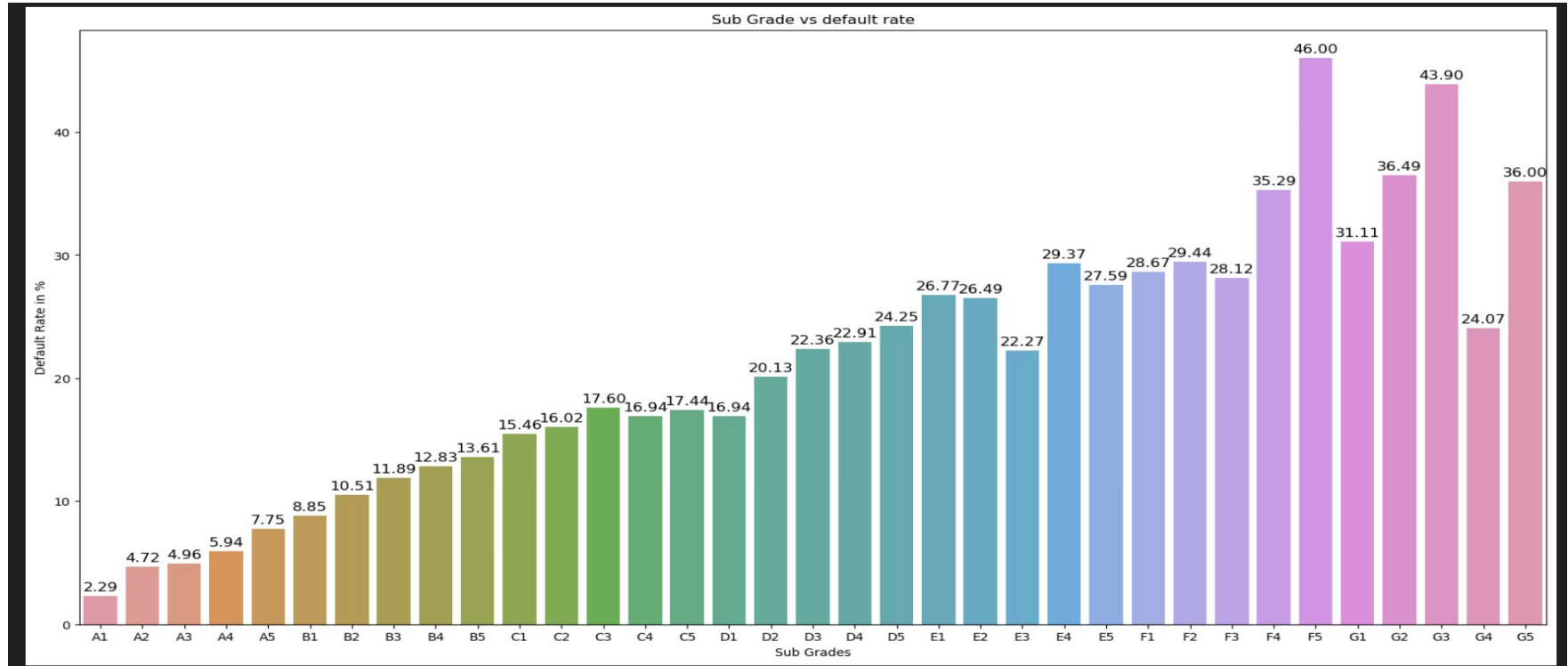
Distribution of Term vs Default rate

Default rate is higher for 60 month term. So the risk is higher for 60 month there is a high change of defaulting for other category home owners



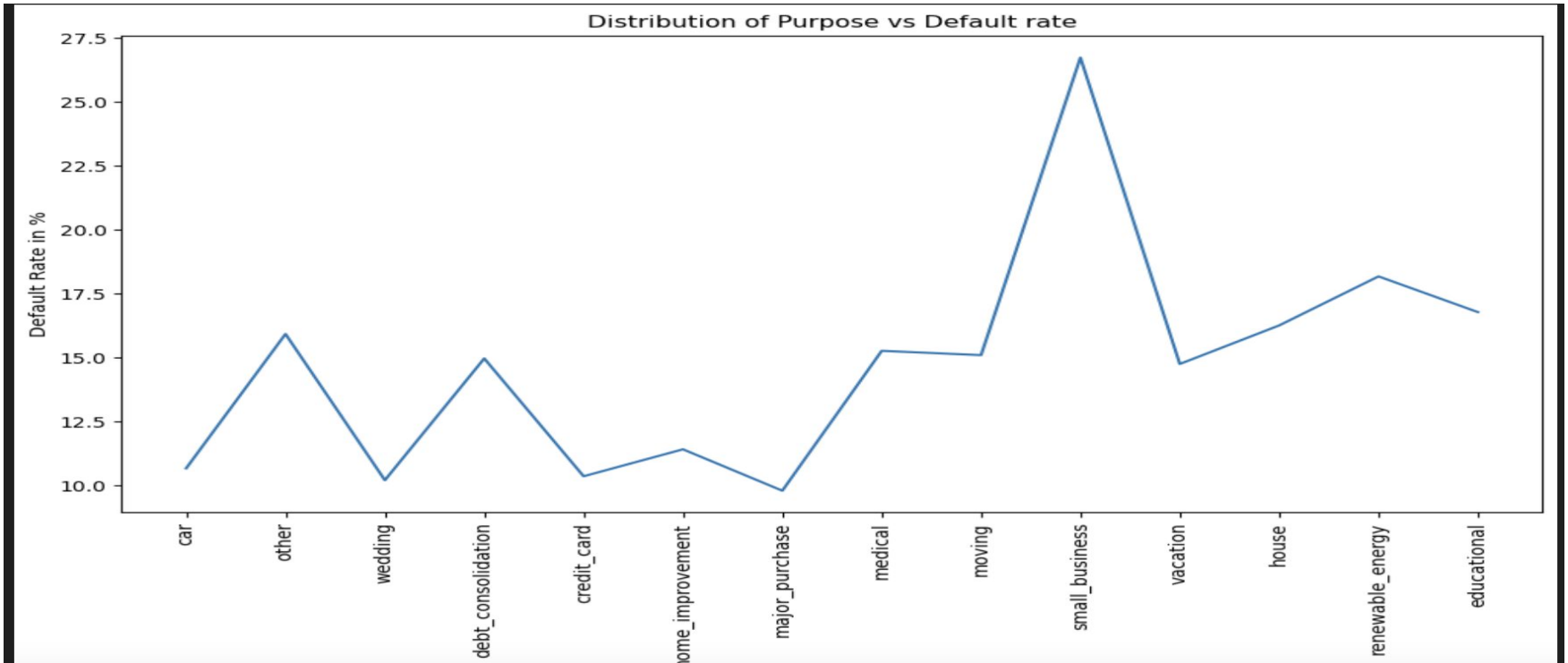
Sub Grade vs Default rate

Sub grades E F and G have higher default ratings so this matches with the distribution of grades vs default rate



Distribution of Purpose vs Default rate

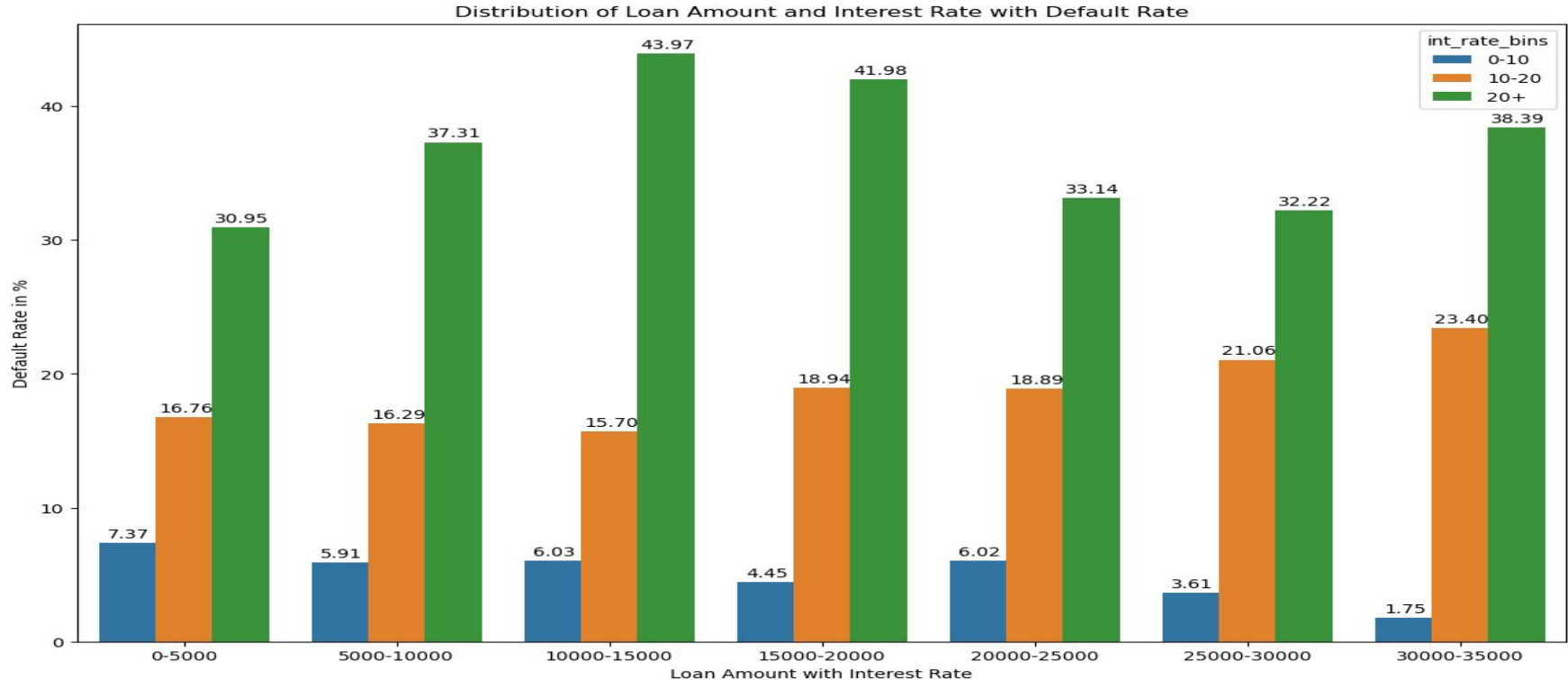
Small business have higher default rate among others and others loan purpose is having almost similar default percentages



Bivariate Analysis

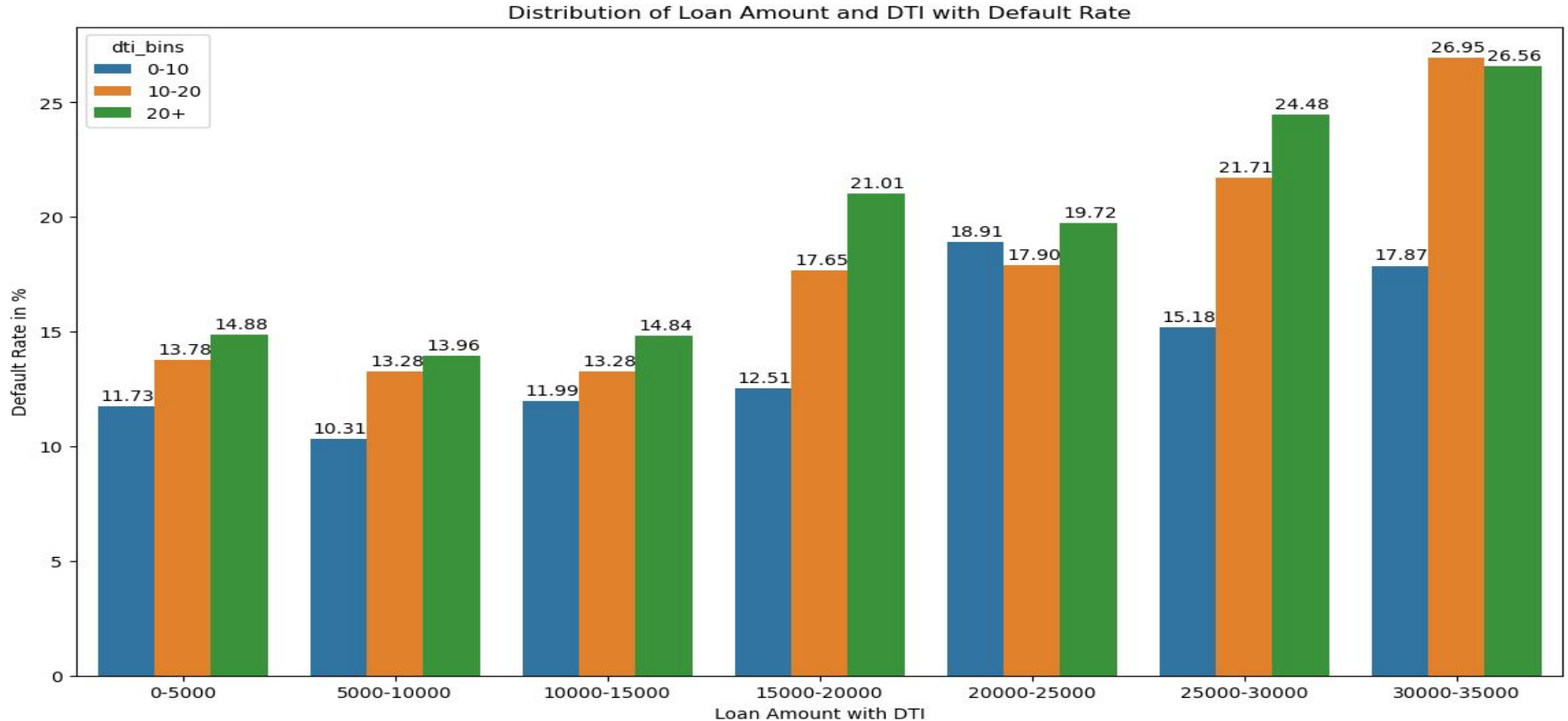
Distribution of Loan Amount and Interest Rate with Default Rate

Loan Amount between 5000 to 20000 with interest rate more than 20% have greater default rates which is likely to have chance of defaulting.



Distribution of Loan Amount and DTI with Default Rate

Higher Loan Amount with DTI has a chance of defaulting.

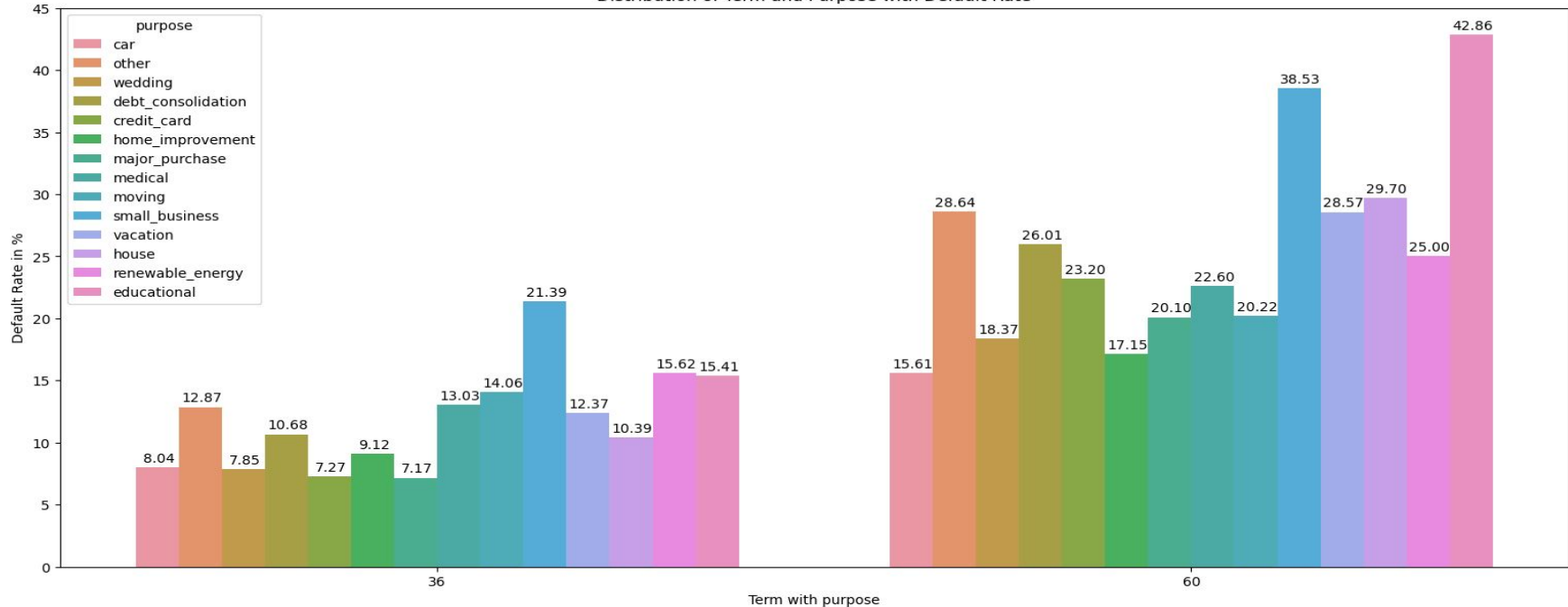


Distribution of Terms and Purpose with Default Rate

For 36 month term, small business is having higher default rating so higher chances of defaulting.

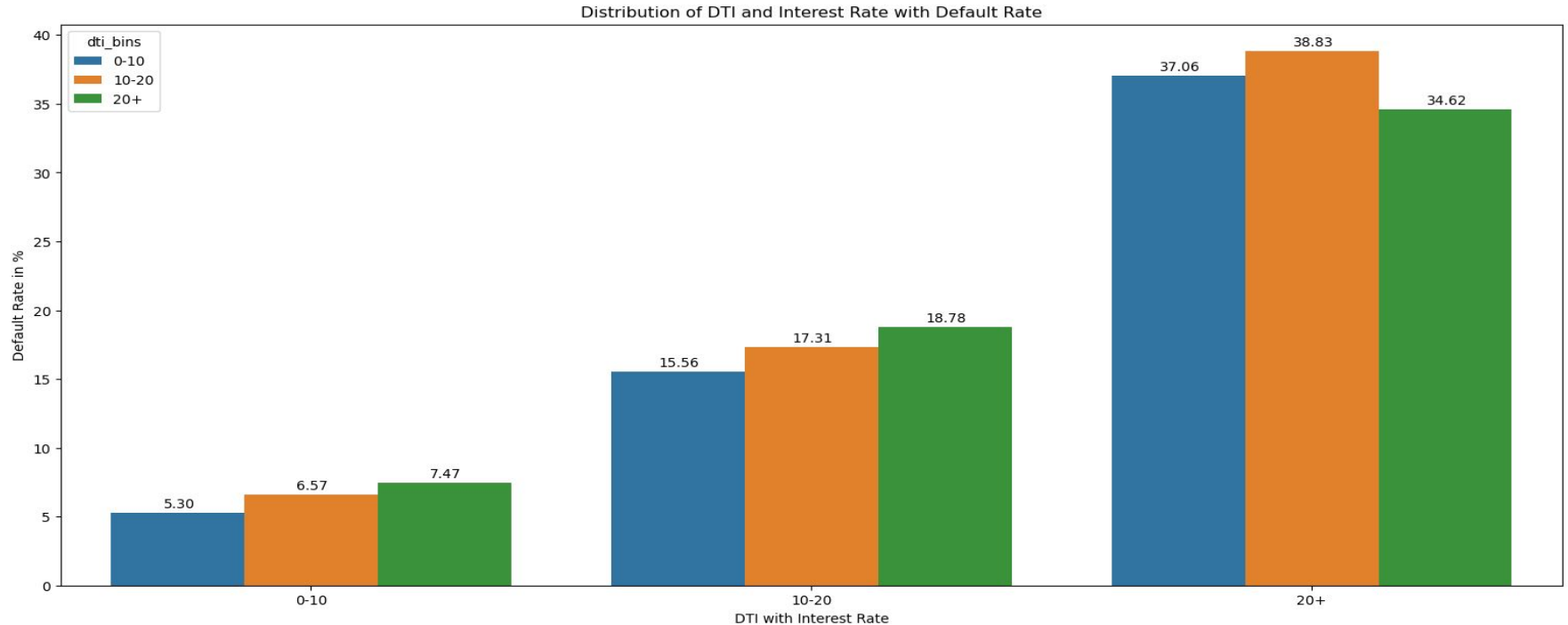
For 60 month term, small business and educational loan is having higher chances of defaulting

Distribution of Term and Purpose with Default Rate



Distribution of DTI and Interest Rate with Default Rate

For lower interest rates (0-10 and 10-20) default rate increases gradually with dti ratio and for interest rate more than 20% default rate and dti ratio does not have a similar pattern



Identify the key driving factors

int_rate_bins

grade

purpose

term

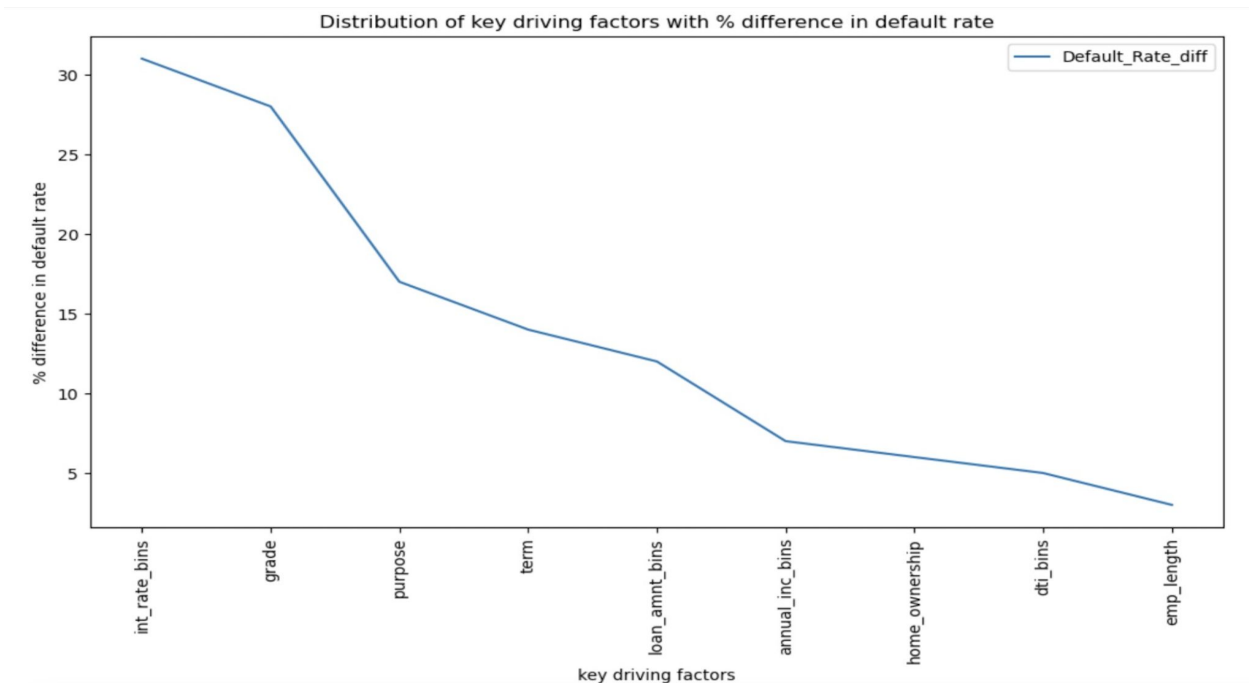
loan_amnt_bins

annual_inc_bins

home_ownership

dti_bins

emp_length



Conclusion

From the analysis we can come to a conclusion that below are the top 6 key driving factors which can be used to analyse the defaulters

1. Interest Rates

2.grade

3.purpose

4.term

5.loan amount

6.annual annual income

.