LENDING CLUB CASE STUDY

SUBMISSION

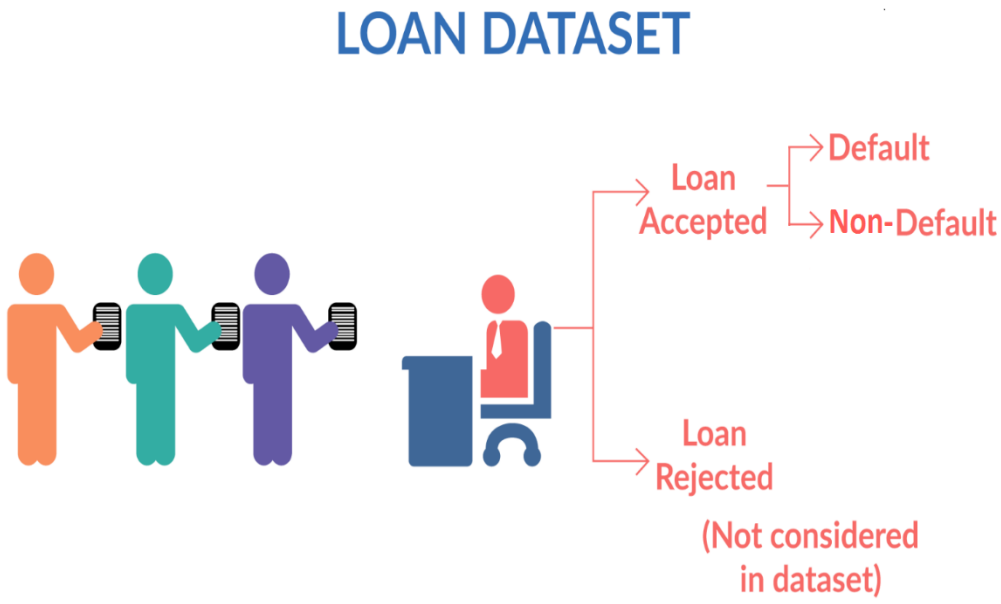
Lending Club Case Study

Problem Statement:

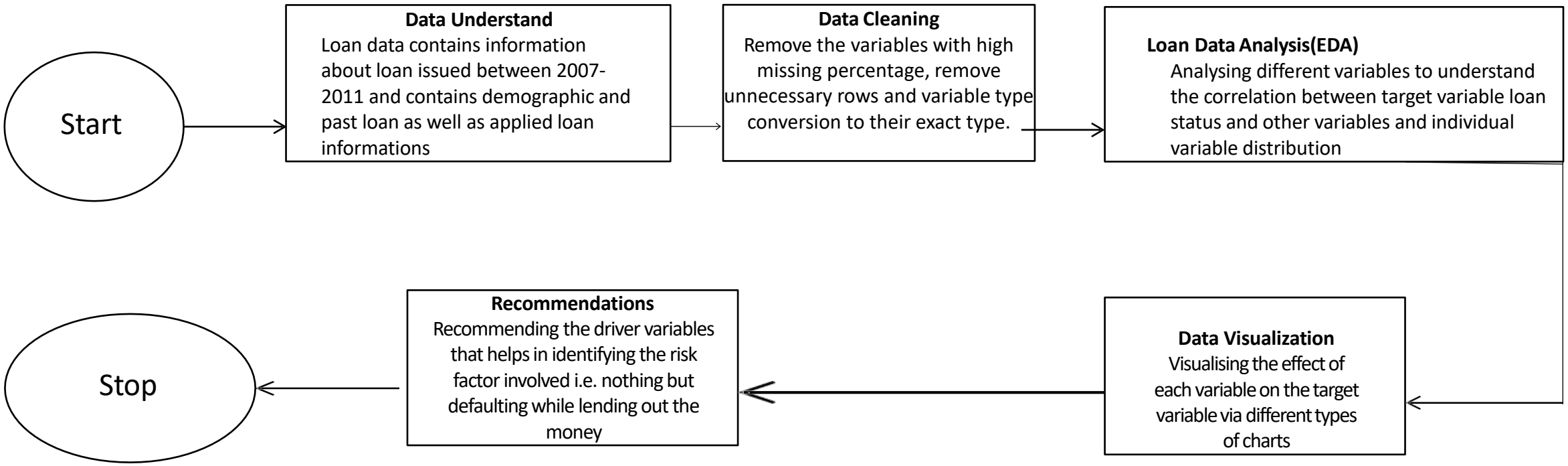
A company who is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures faces two types of risk:

- Applicants who are likely to repay the loan, not approving the loan for them results in a loss of business to the company
- Applicants who are not likely to repay the loan, i.e. he/she is likely to default, approving the loan for them leads to a financial loss for the company

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.

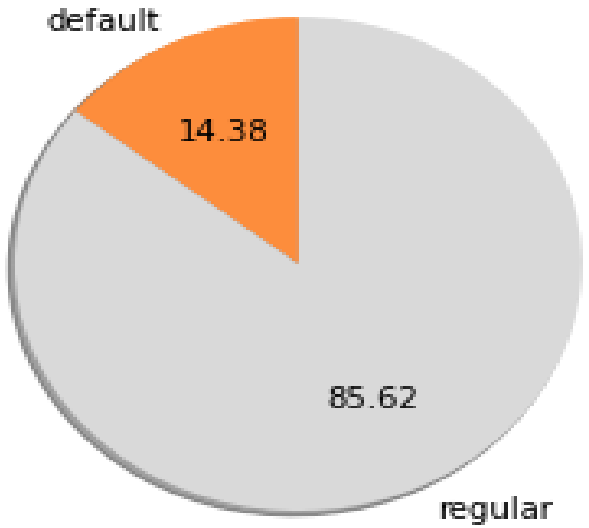


Problem solving methodology



Analysis

This analysis starts with cleaning of the data where out of the 111 columns and 39717 rows, 27 columns and 38577 rows were left. The reason behind dropping includes null values and unnecessary data and irrelevant variables for the analysis.



This case study was mainly divided into 4 categories.

1. Data Cleaning
2. Univariate Analysis
3. Bivariate Analysis
4. Multivariate Analysis

The flow of the analysis is as follows:

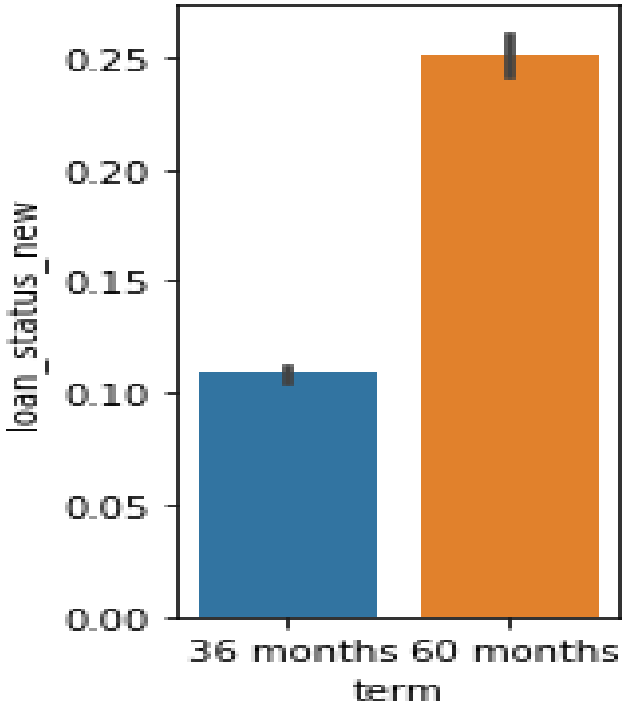
The data was cleaned first.

Then univariate analysis was done for variables and business insights were obtained from them. Like for the entire data ~ 14% was default rate. And which variables showed the clear trend of increasing default rate for each group or bin of the data.

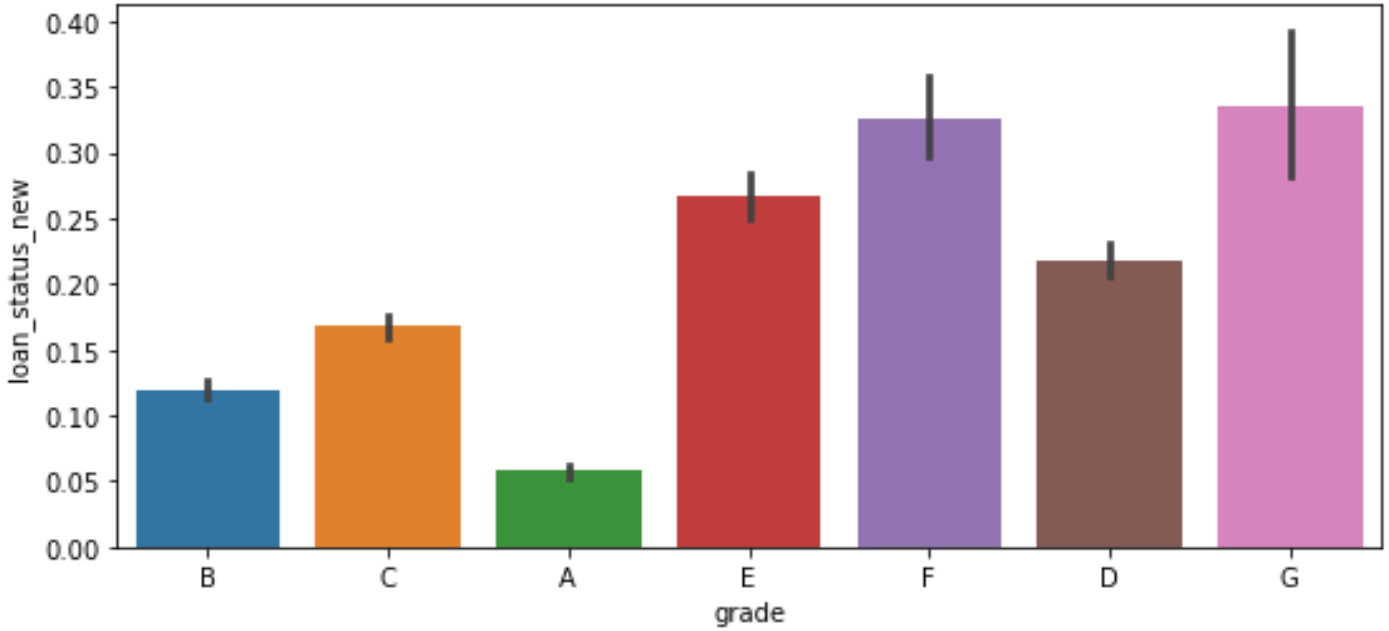
Then bivariate analysis was done and effect of two variables as well as effect of one variable nested in the data segmented by another variable on risk of default is also observed.

Finally, combined effect of all the variables on default rate is also done under multivariate analysis.

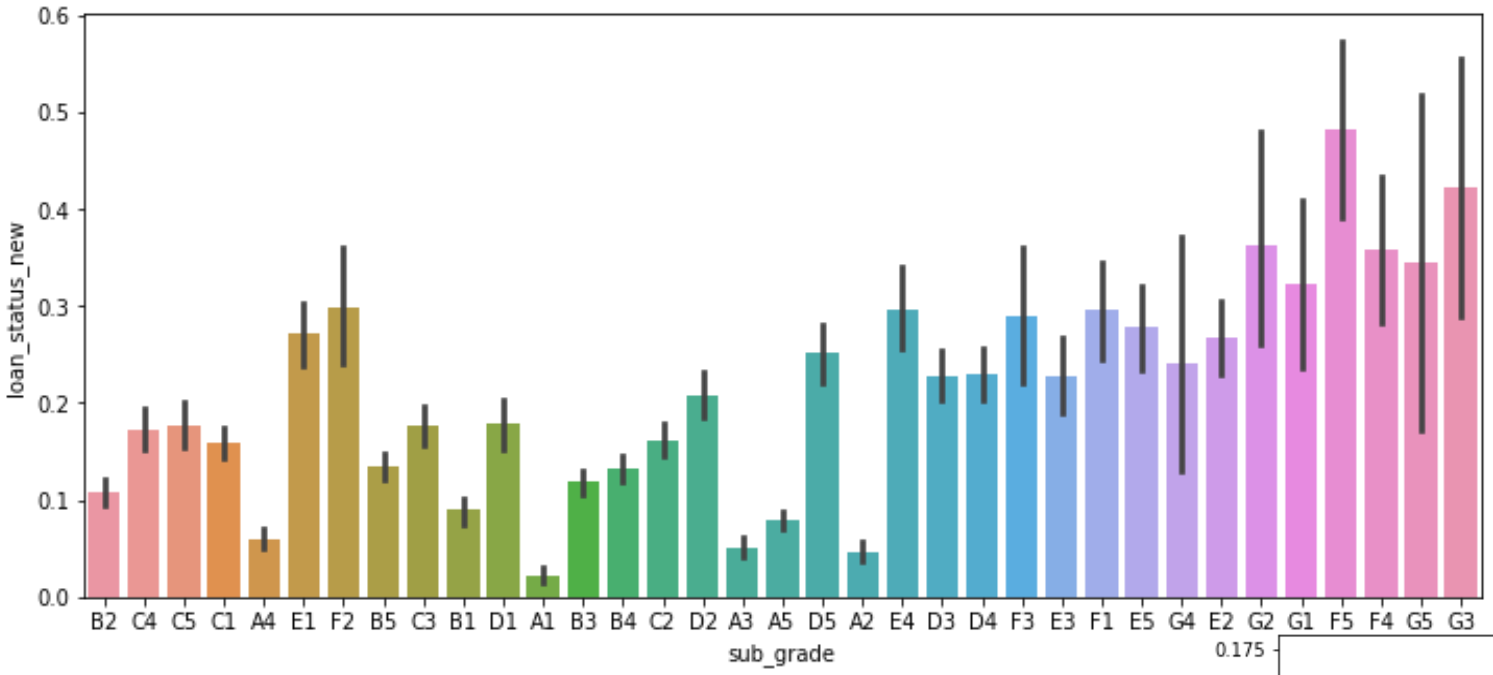
Analysis for loan term, grade variables



- Loans with longer loan term have higher default rate than for shorter loan term
- Default rate increases with the LC assigned grade for the loan.

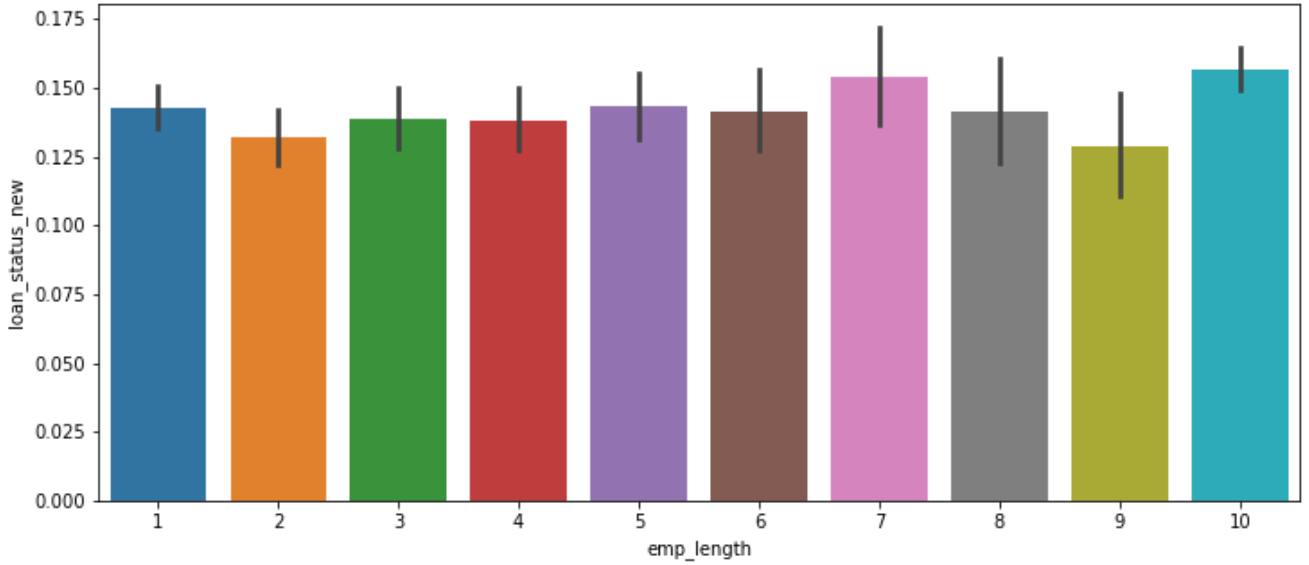


Analysis for subgrade and emp_length variables

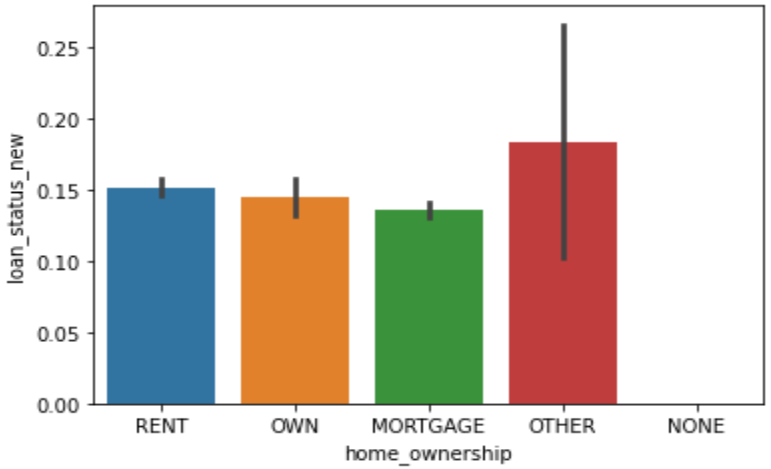


- With sub grade, their is break in the trend but over all it follows the trend i.e. higher the subgrade when alphabetically ordered, higher is the default rate

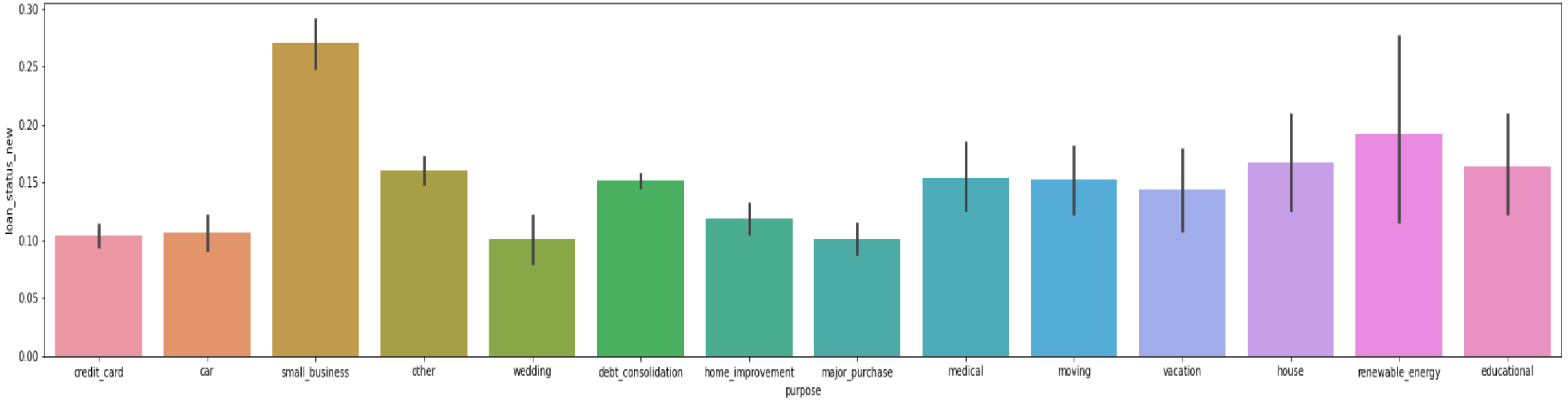
- Length of employment is not a good explanatory variable for default rate as it gives a flat trend with increase in employment length for default rate.



Analysis for type of home ownership and purpose of loan application



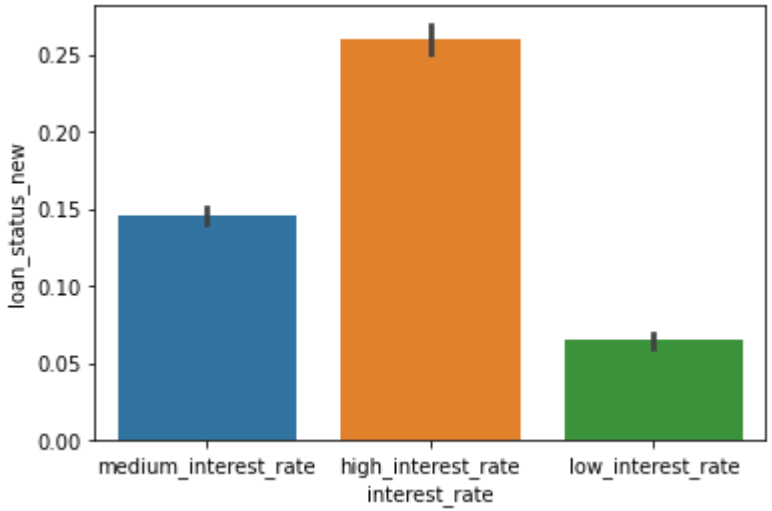
- Home ownership of other type is riskiest profile to lend which is because who does not have their fixed home, they default the most.
- The loan application which is for the purpose small_business has the highest default rate because sometimes business may not work out and hence is the riskiest profile for lending.

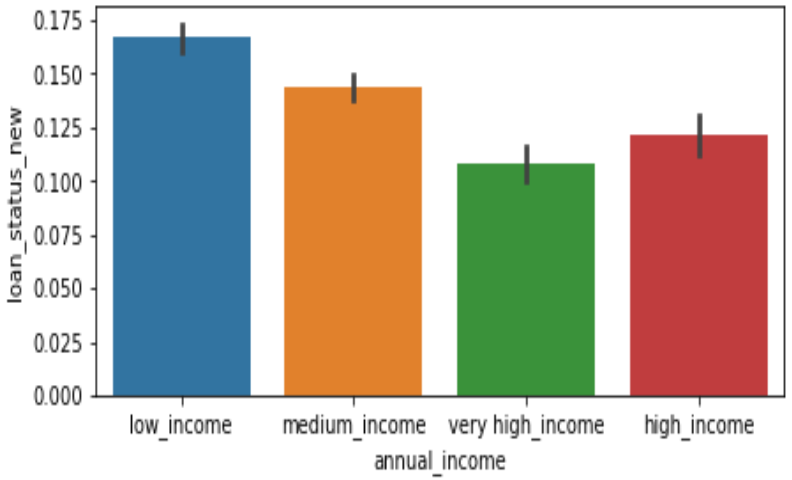


Descriptive Analysis for all the numeric variables in the loan data set for understanding data quality

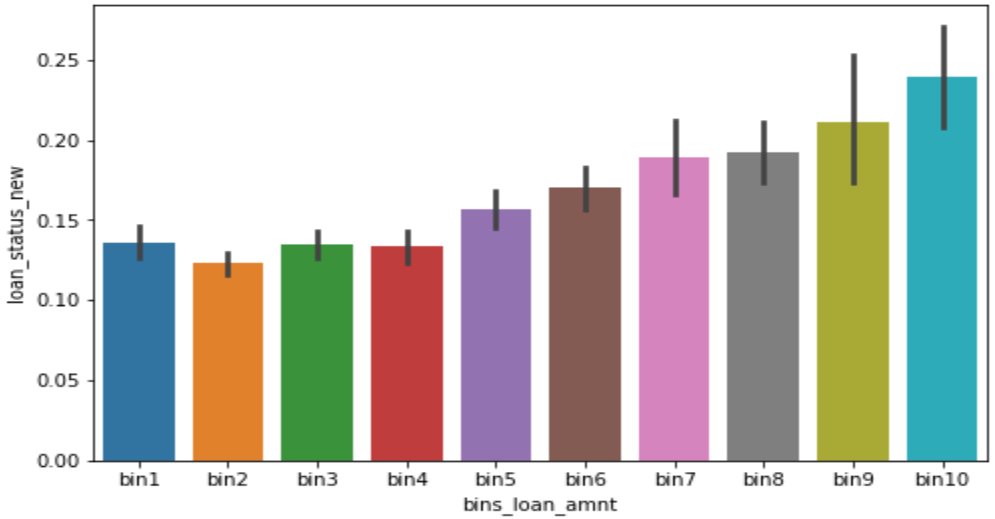
	loan_amnt	funded_amnt	funded_amnt_inv	int_rate	installment	annual_inc	dti
count	37544.000000	37544.000000	37544.000000	37544.000000	37544.000000	3.754400e+04	37544.000000
mean	11119.329986	10853.560489	10280.495739	11.963319	324.650164	6.940708e+04	13.284186
std	7354.098954	7096.316776	7034.124211	3.683012	208.901055	6.467698e+04	6.660551
min	500.000000	500.000000	0.000000	5.420000	15.690000	4.000000e+03	0.000000
25%	5500.000000	5400.000000	5000.000000	8.940000	167.370000	4.100000e+04	8.160000
50%	10000.000000	9600.000000	8850.000000	11.830000	280.000000	6.000000e+04	13.380000
75%	15000.000000	15000.000000	14075.000000	14.420000	428.942500	8.300000e+04	18.550000
max	35000.000000	35000.000000	35000.000000	24.400000	1305.190000	6.000000e+06	29.990000

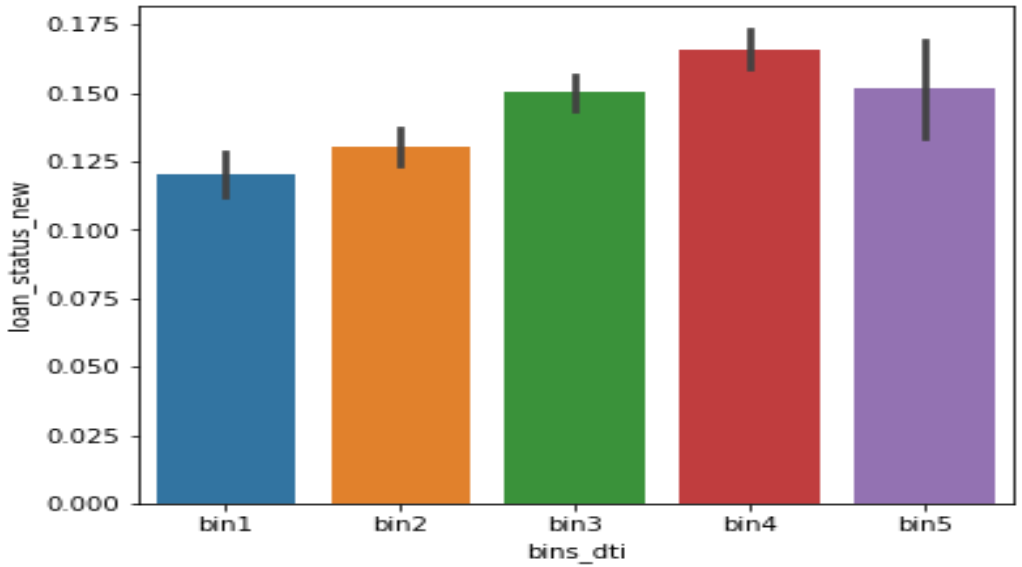
Analysis for Binned variables like interest rate, loan_amount, annual_income and debt to income ratio





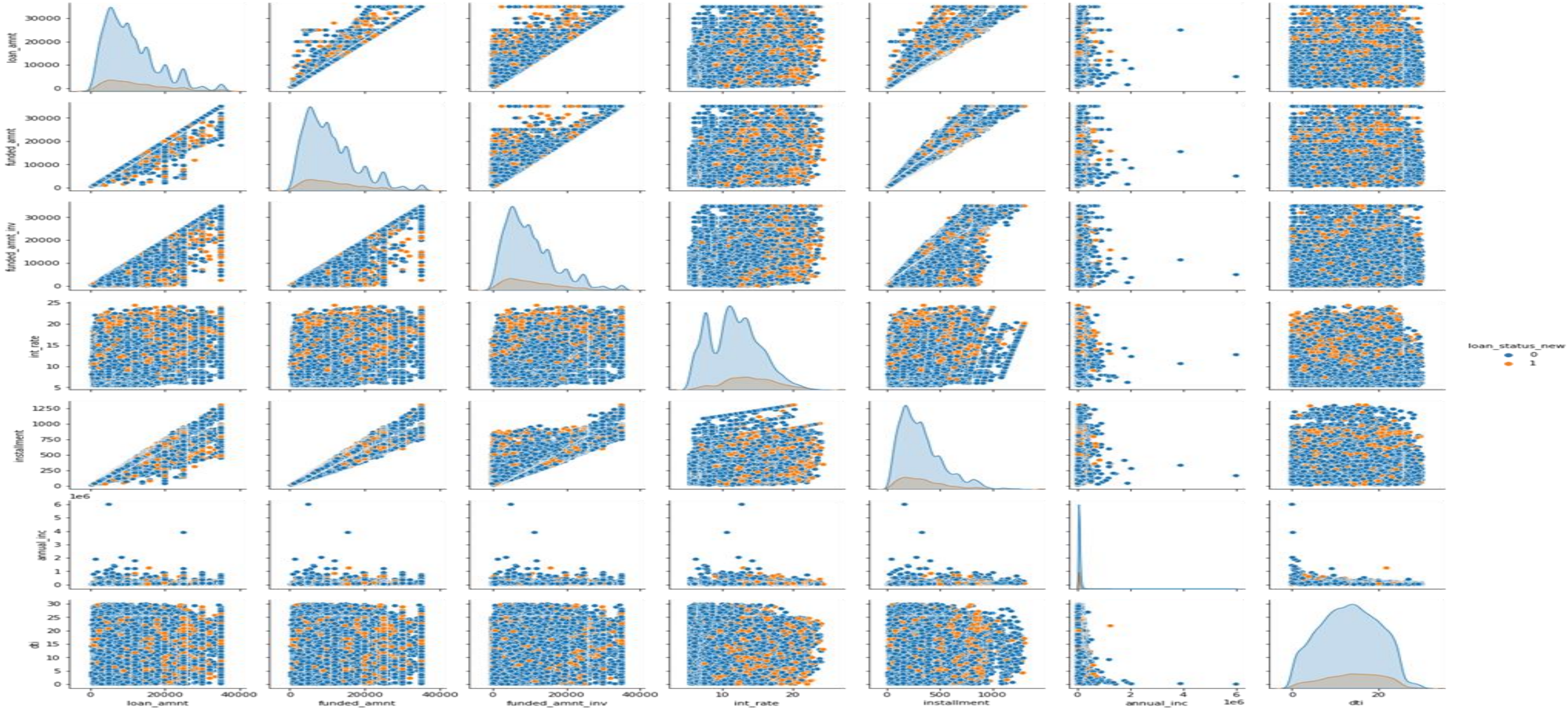
- Higher the interest rate , higher is the default rate and the trend is very clear.
- Higher the loan_amount , higher is the default rate and the trend is very clear except the first bin i.e very small ticket size loan.
- Higher the annual income, lower is the default rate and the trend is very clear which means a person with higher income can smoothly payback the loan without defaulting.
- Higher the debt to income ratio , higher is the default rate and the trend is very clear except the last bin.





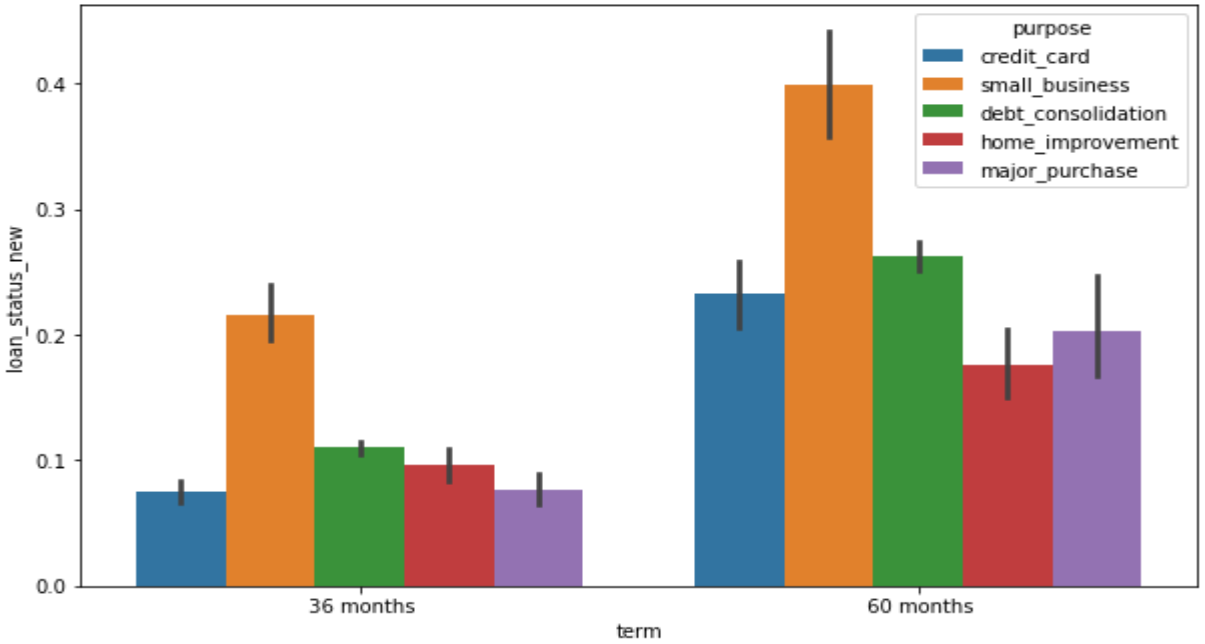
Bivariate analysis for all the numeric variables

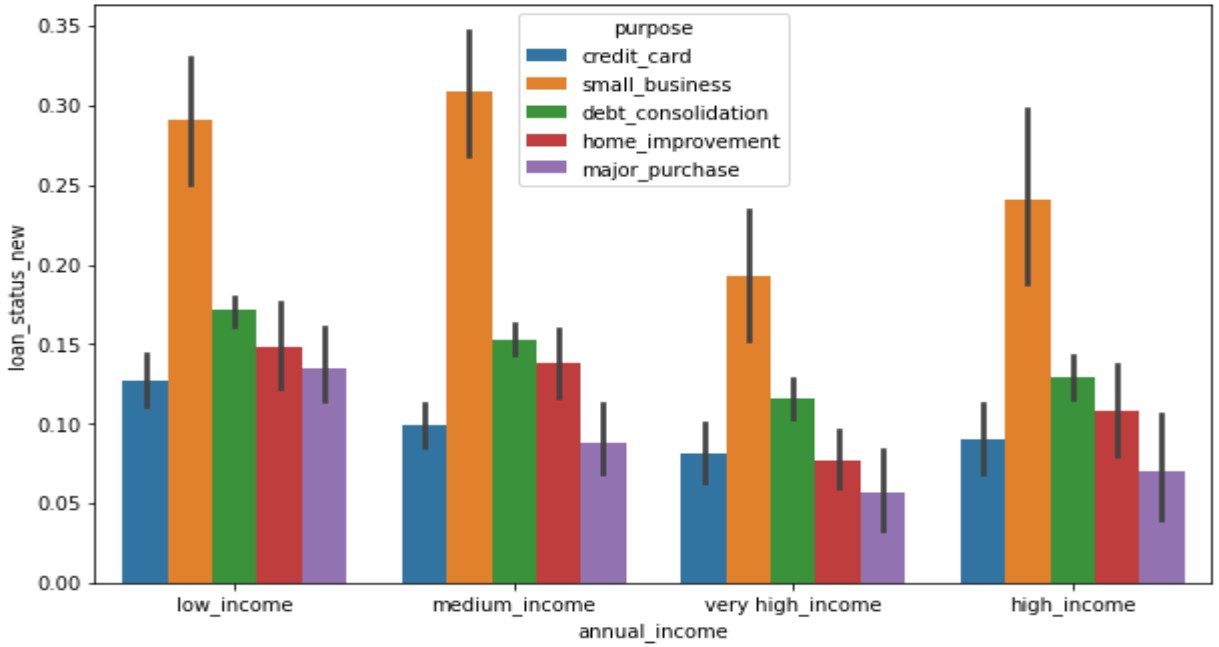
Higher the loan amount and higher the interest rate, higher default and similar other observations



Top 5 loan type wise segmented bivariate analysis

Top 5 most popular loan types are: debt_consolidation, credit_card, home_improvement, major_purchase, small_business



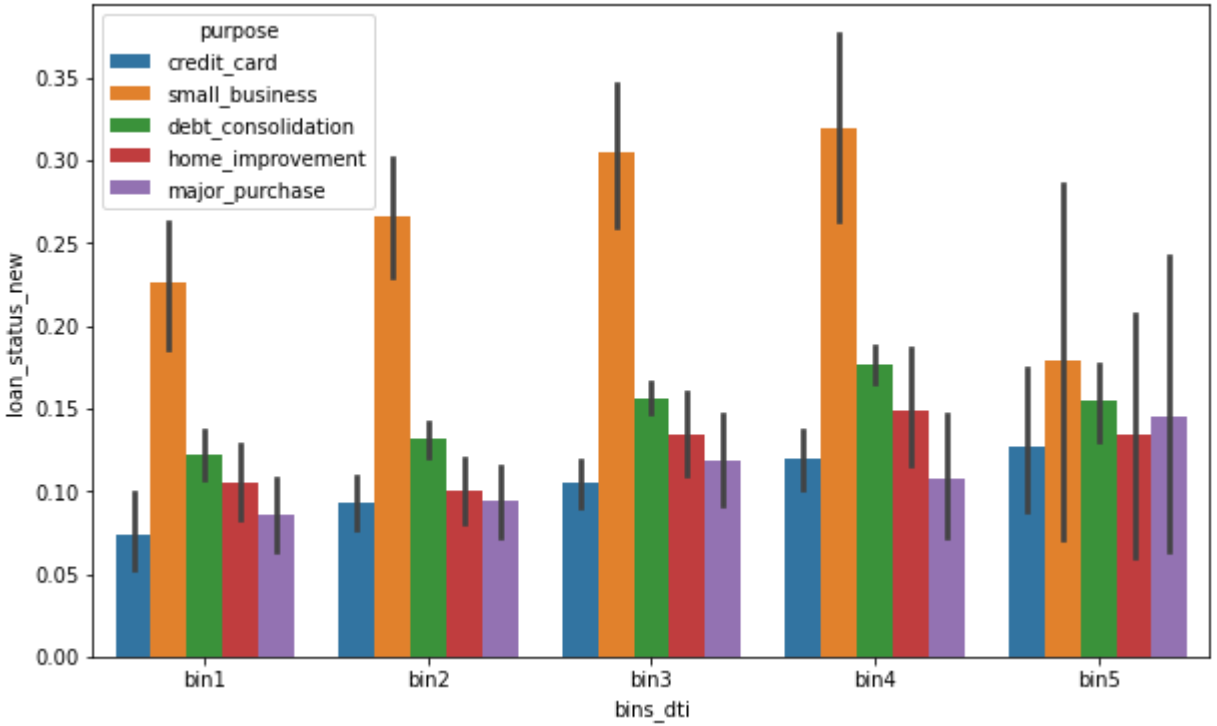


For the above mentioned 5 loan types, on average 60 months loan term shows higher risk profile. Also, Small business consistently shows higher risk trend in both the loan term segments, but the risk profile for home_improvement loan type and major purchases are different across the two different loan terms.

For the above mentioned 5 loan types, on average Small business consistently shows higher risk trend in all the income buckets and the default rate trend across each income bucket for the loan types are same.

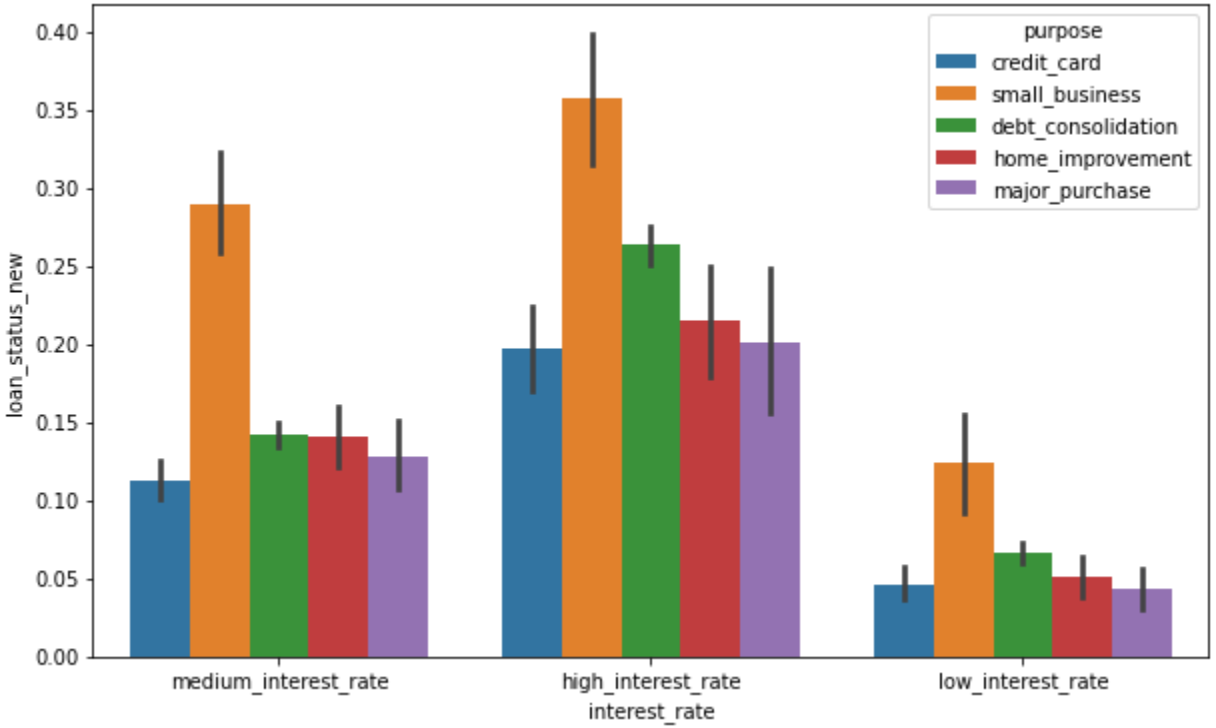
Top 5 loan type wise segmented bivariate analysis

Top 5 most popular loan types are: debt_consolidation, credit_card, home_improvement, major_purchase, small_business



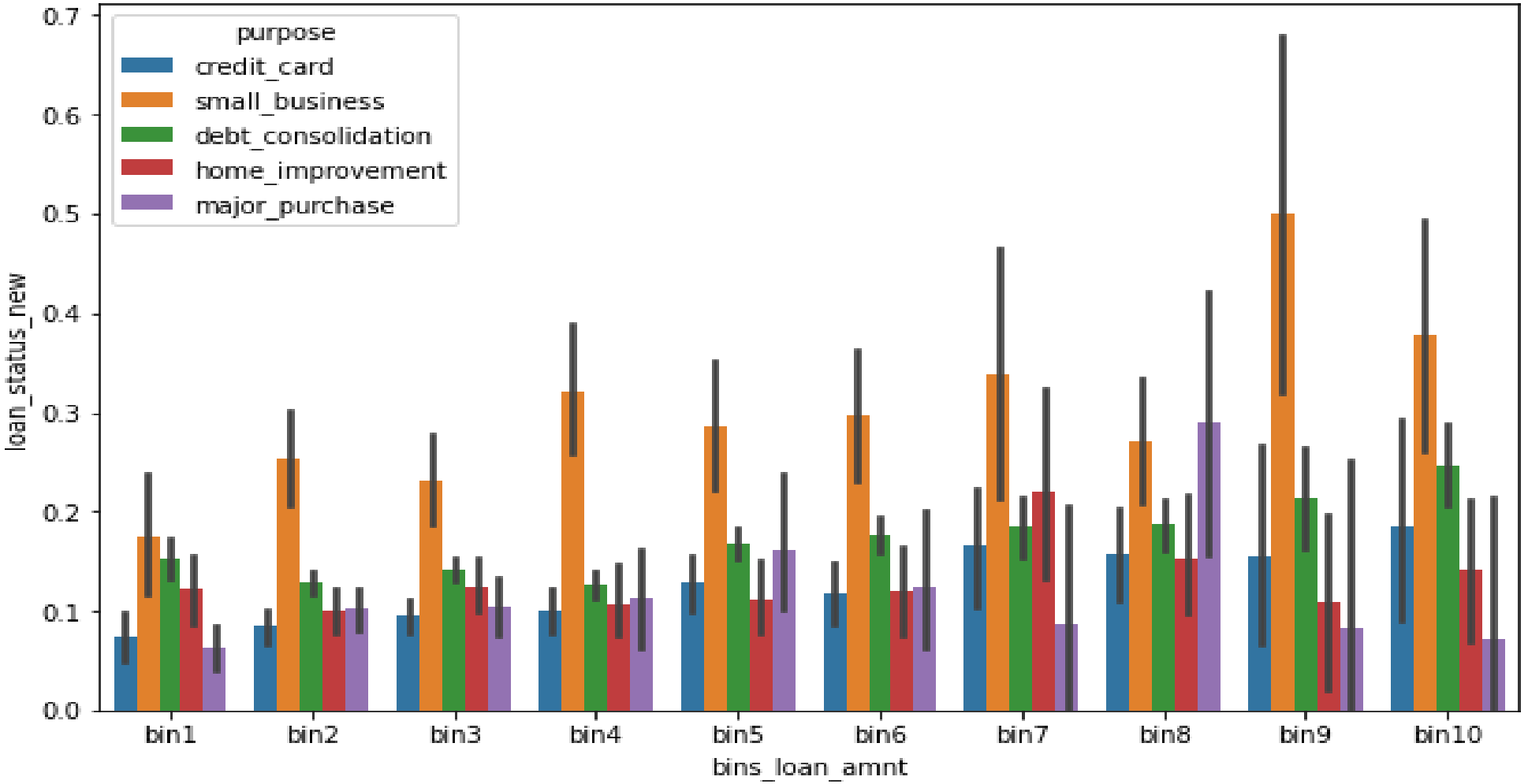
For the above mentioned 5 loan types, on average Small business consistently shows higher risk trend in all the debt to income buckets where bin1 has lowest dti and bin 5 has highest dti. and the default rate trend across each debt to income bucket for the loan types are same.

For the above mentioned 5 loan types, on average Small business consistently shows higher risk trend in all the interest rate buckets and the default rate trend across each interest rate bucket for the loan types are same.



Top 5 loan type wise segmented bivariate analysis

Top 5 most popular loan types are: debt_consolidation, credit_card, home_improvement, major_purchase, small_business



For the above mentioned 5 loan types, on average Small business consistently shows higher risk trend in all the loan_amount buckets where bin1 has lowest disbursed loan amount and bin 10 has highest lowest disbursed loan and the default rate trend across each debt to income bucket for the loan types are same.

Recommendations

No	Variable	Recommendations
1	Interest rate	Lower interest rate is better than higher interest rate because higher the interest rate , higher is the default rate.
2	Loan amount	Lower loan amount is better than higher interest rate because higher the loan amount , higher is the default rate.
3	Loan purpose	Small_business loan are very risky and is associated with high default rate. Hence bank should avoid disbursing such loans. Should focus on credit_card associated with lower risk profile.
4	Loan Grade	Grades A to C are associated with low risk and hence can be used. Loan with grade D and above should be avoided.
5	Debt to Income	Anything less than 12 dti, is still safe but above 12 is very risky. Hence bank should not give loan over 12 dti.
6	Home Ownership	People who are home owner are associated with lowest risk where as people with other type is the riskiest. Also people with rent mortgage are risky as well hence bank should give loan to only those who have a own house.