Team

Himanshu Pal

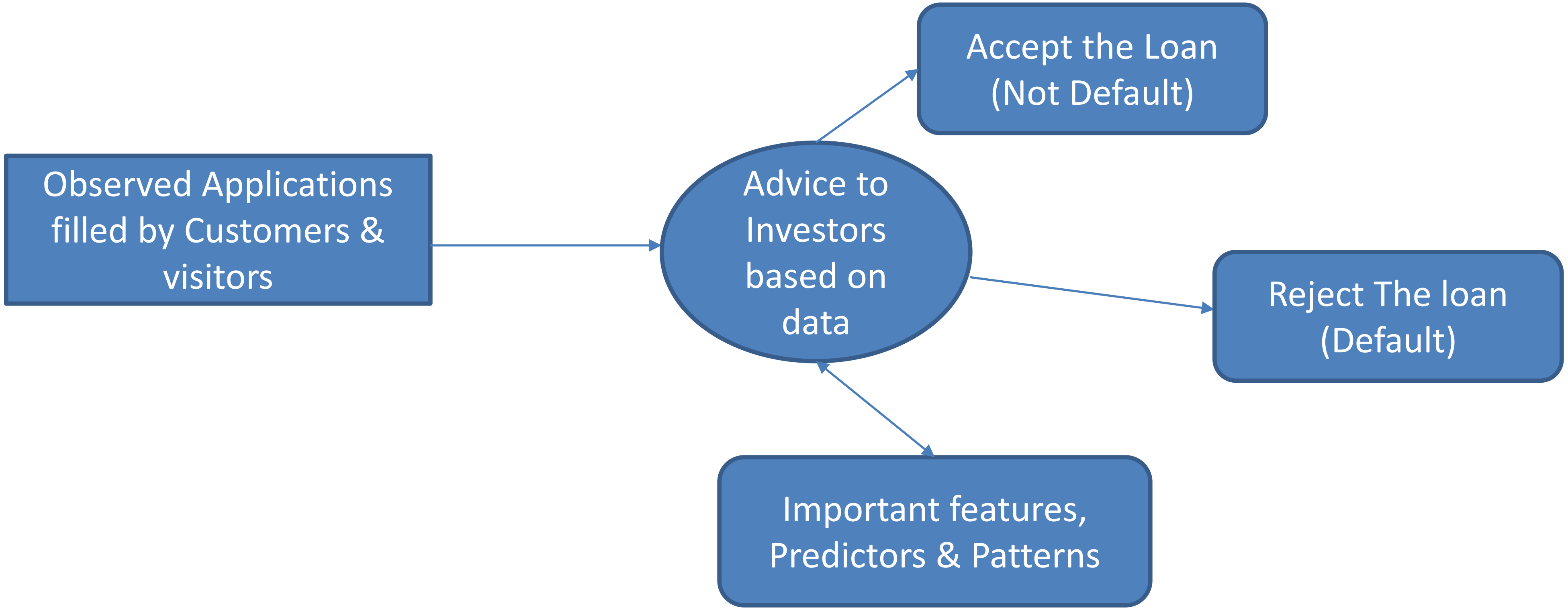
Manish Kaushal

LENDING CLUB CASE STUDY

Business Objective

Lending Club 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 online interface.

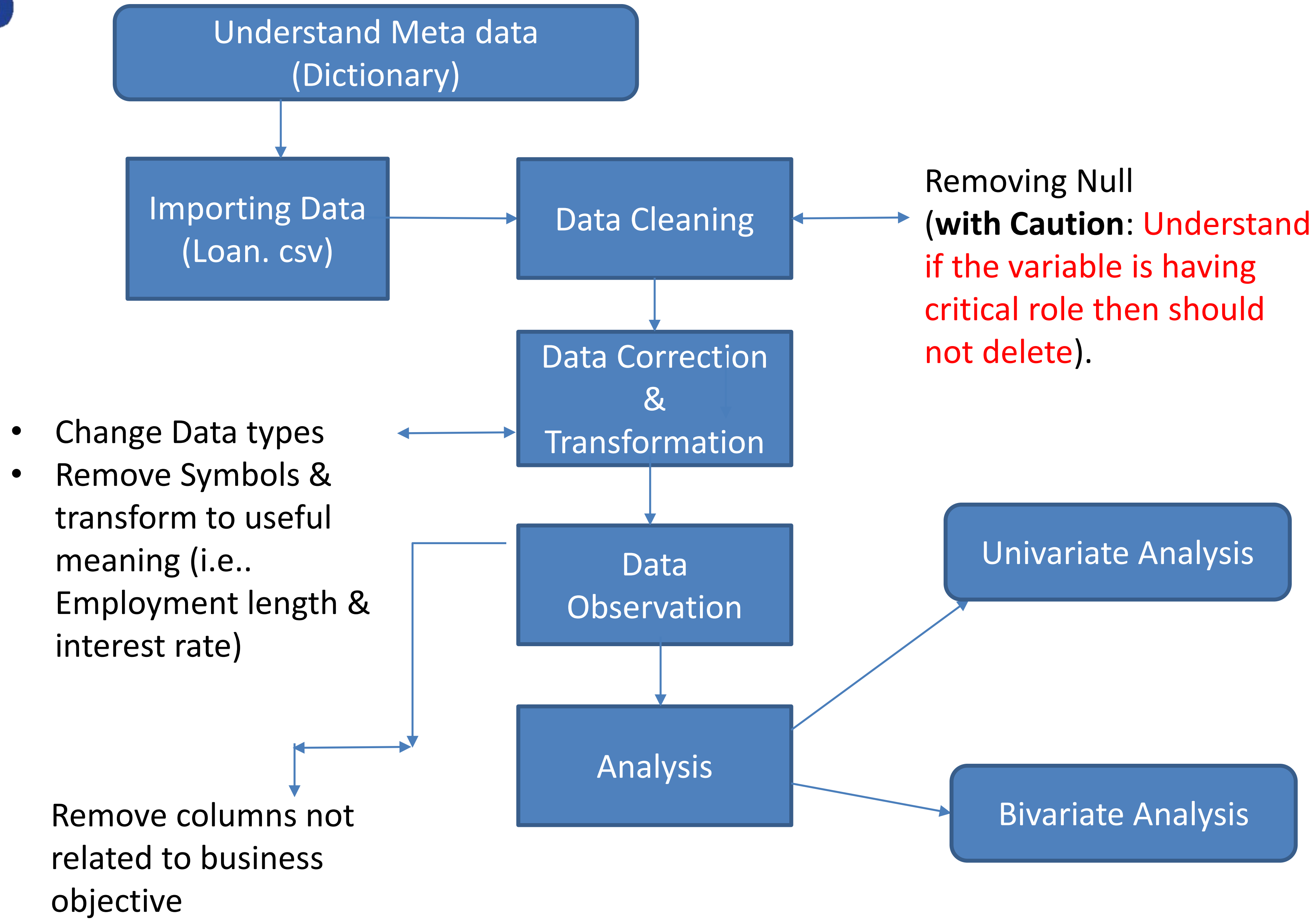
loan borrowers who 'default' cause to the large amount loss to the lenders. The company & investors thus wants to understand the driving factors behind loan default, i.e. the variables which are strong indicators of default.



Based on the advice by lending club, investors decide to invest or lent money to borrower

There are three important variables mentioned below which we need to understand:

- Loan Amount - Amount of money that Borrower asked.
- Funded Amount - Amount of money lending club suggested.
- Funded Amount Invest - Amount of money that actually lent to customer.


Understand Meta data (Dictionary)
Importing Data (Loan. csv)
Data Cleaning
Removing Null (**with Caution**: Understand if the variable is having critical role then should not delete).
Data Correction & Transformation
• Change Data types
• Remove Symbols & transform to useful meaning (i.e.. Employment length & interest rate)
Remove columns not related to business objective
Data Observation
Analysis
Univariate Analysis
Bivariate Analysis

The data set while upload the “loan.csv” file we have 111 attributes/columns.

The number of relevant attributes were brought down to 32 after performing the following processes:

- Dropping columns with all NULLs.
- Dropping rows with high percent of missing values from the columns.
- Dropping columns that represented the customer behavioral variables like- delinq_2yrs, open_acc, pub_rec, etc.

There are broadly three types of variables -

- **Demographical variables** – these are related to the applicant like age , occupation, employment length ,etc.
- **Loan characteristics** - such as amount of loan , interest rate , purpose of loan , etc.
- **Customer Behavior attributes** – such as delinq_2yrs, open_acc, pub_rec, etc. These are generated after the loan has been approved .

Now , customer behavior variables are not available at the time of loan application and thus they can not be used as a predictor for credit approval . Also loan status as ‘current ’ are not either fully paid or charged of so we can drop them.

Uni-Variate Analysis

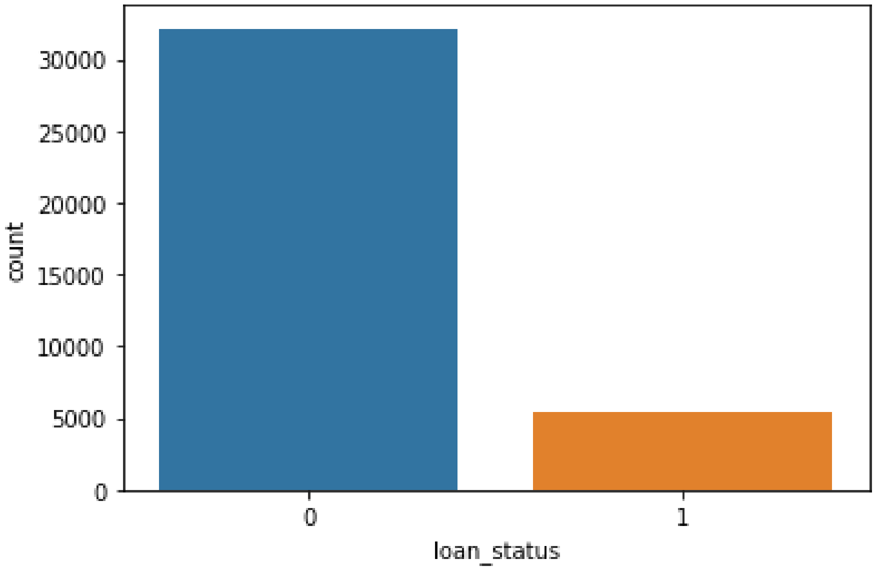
For univariate analysis we can proceed by following ways:

- Default rate across various categorical features - We can plot default rate vs various attributes like term, grade ,sub_grade, etc. This will visually show how they affect each other.
- For continuous features , we can preform binning and then perform univariate analysis. We can divide the continuous variables data into bins and give them values like - high , medium ,low and plot them against default rates.

Following are important columns/features on which can be used for analysis:

loan_amnt,term,int_rate,sub_grade,grade,annual_inc,verification_status,purpose,dti,funded_amnt,funded_amnt_inv,purpose etc.

Loan Status vs count

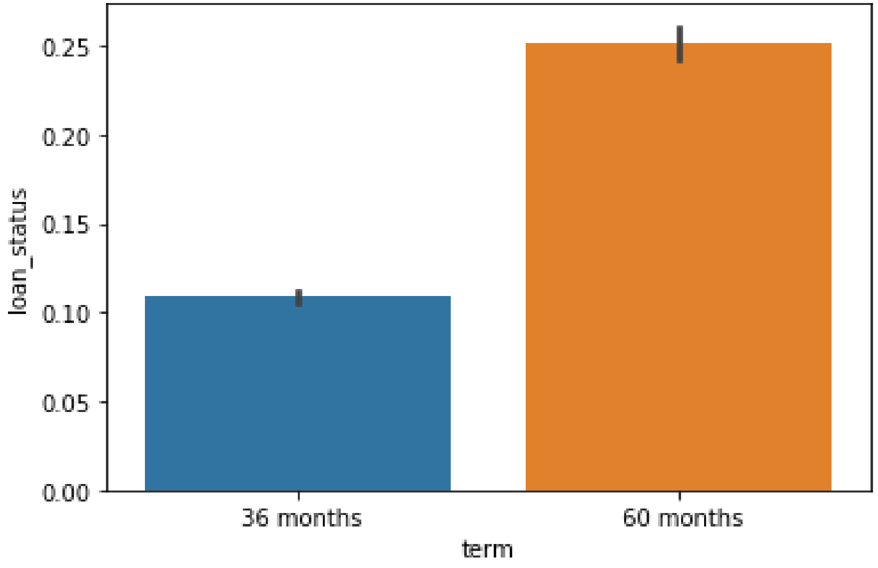


Inferences

- '0' is for non-defaulters
- '1' is for defaulters .

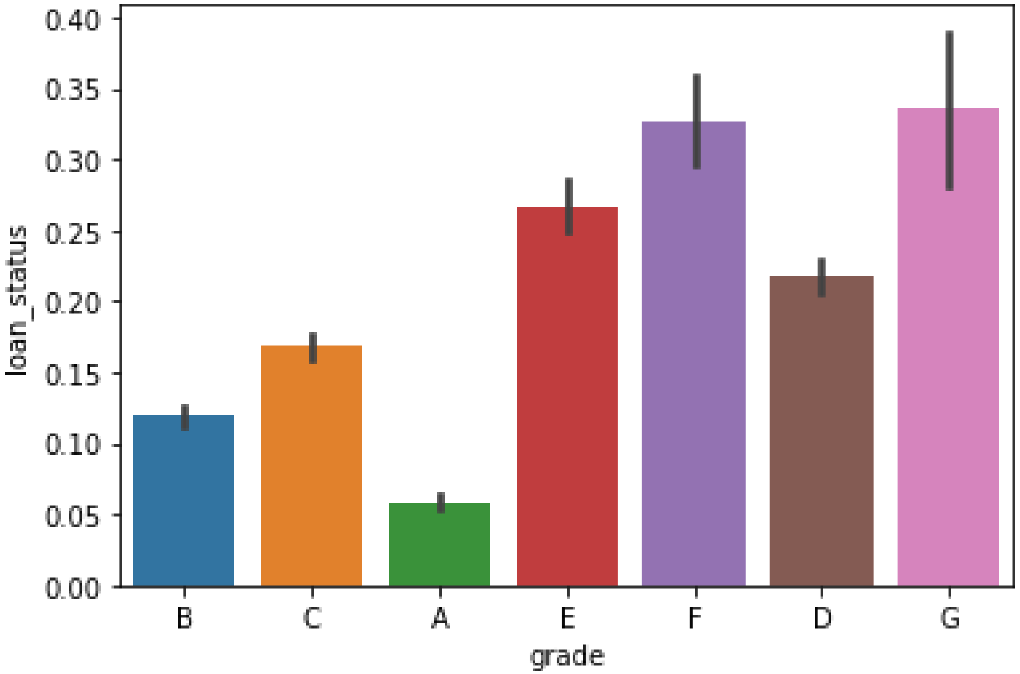
Observations: more than 300,000 borrowers paid back to the investor & nearly 5000 defaulted
Key pointer: It's a huge difference between defaulters & non defaulters.
Overall default rate is 14%

Loan Terms vs Default Rate



People who took loan for a term of 60 months are more likely to default as compared to those for 36 months.

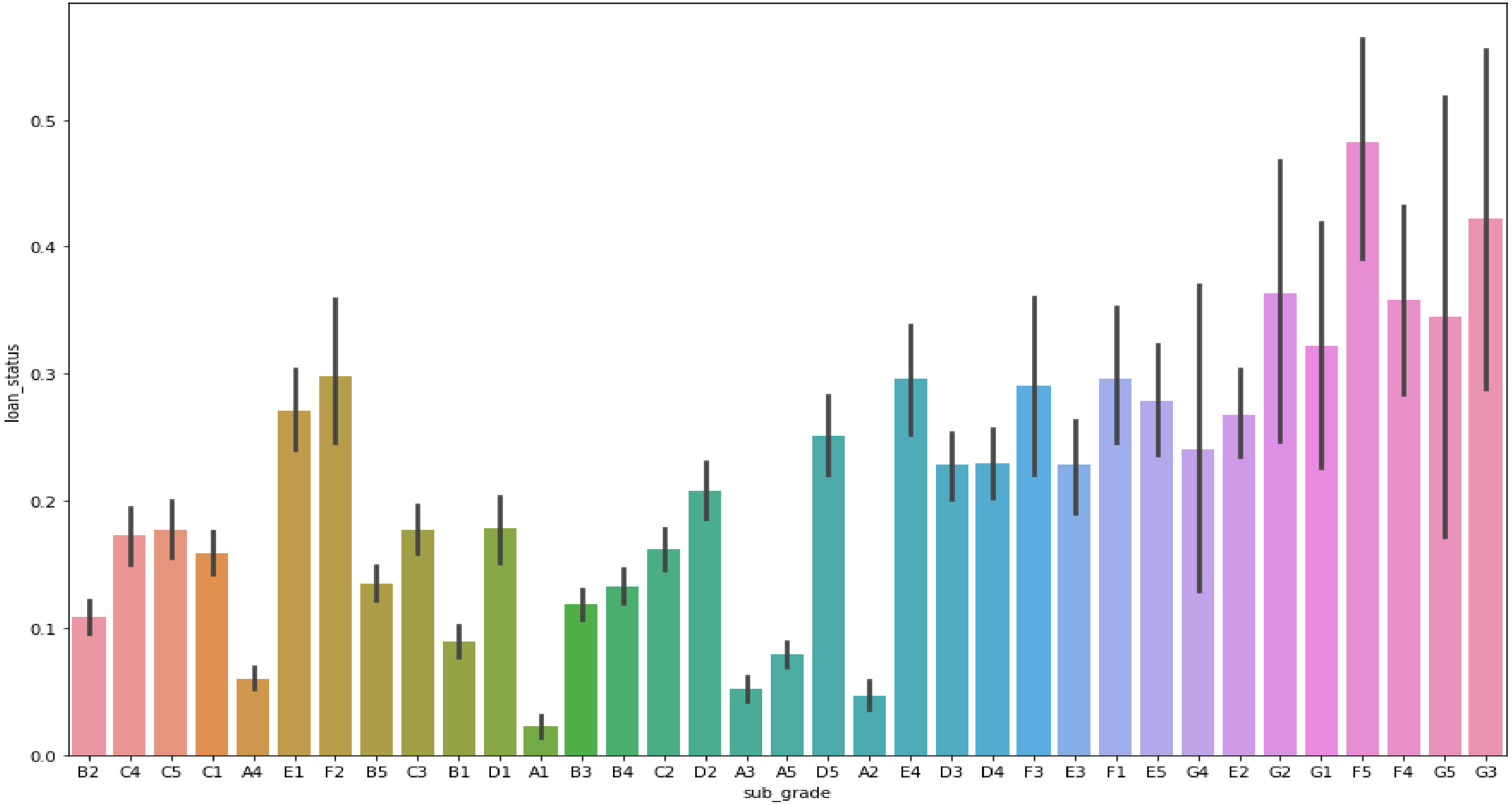
Grades vs Default Rate



Inferences

The graph clearly shows that A grade have less default percentage as compared to B and so on for higher grades . The grade assigned by lending club is one of the factors that is helpful for investors to approve the loan or reject the loan.

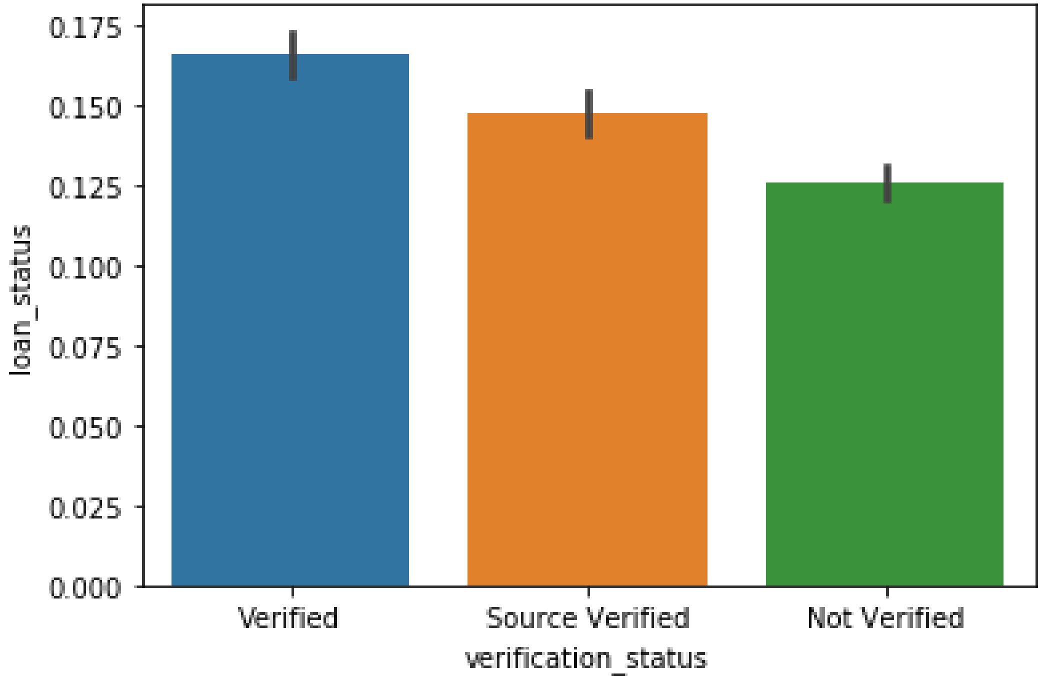
Sub Grades vs Default Rate



Inferences

As the previous plot , sub-grade also behaves on same lines . A1 grade is less risky to be given the loan as compared to A2 or A3 . This will be helpful for investors to approved loan or should be rejected.

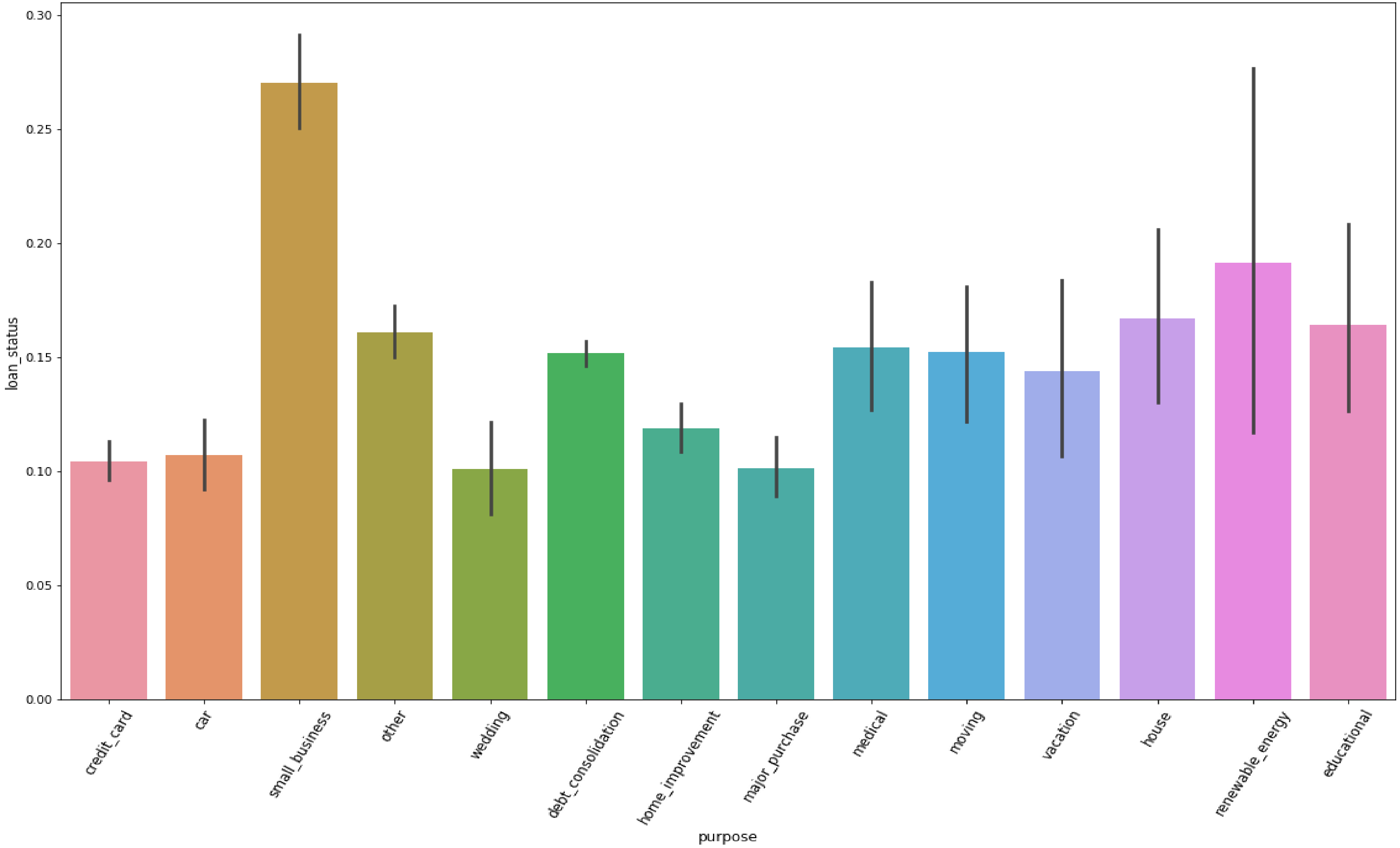
Verification Status vs Default Rate



Inferences

This is quite interesting outcome of analysis , that borrowers with verified income source by Lending Club default more as compared to not verified ones.

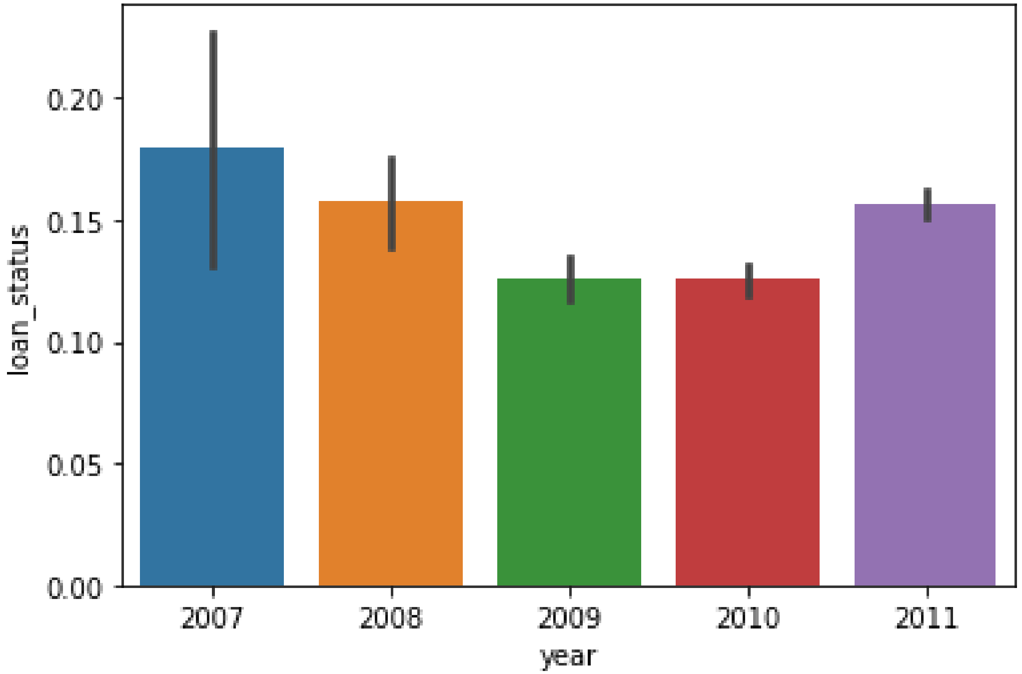
Purpose vs Default Rate



Inferences

The plot clearly shows that: small business default most , then renewable energy are second most defaulters and so on. Whereas, Wedding and Major Purchase are among the good options for granting the loans.

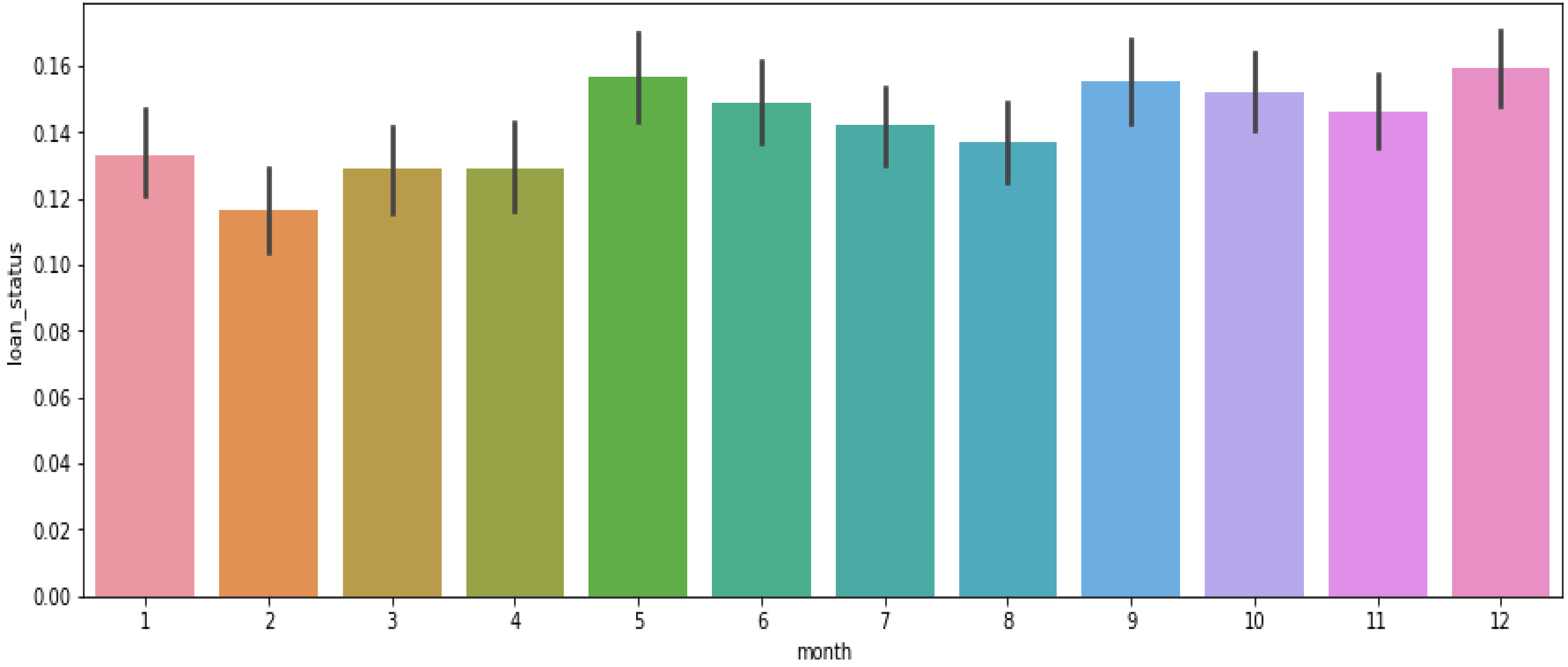
Years vs Default Rate



Inferences

- The default rate had shown abrupt increase in 2011, which shown decline since year 2008 to year 2010.
- In year 2007 default rate was maximum.

Month vs Default Rate

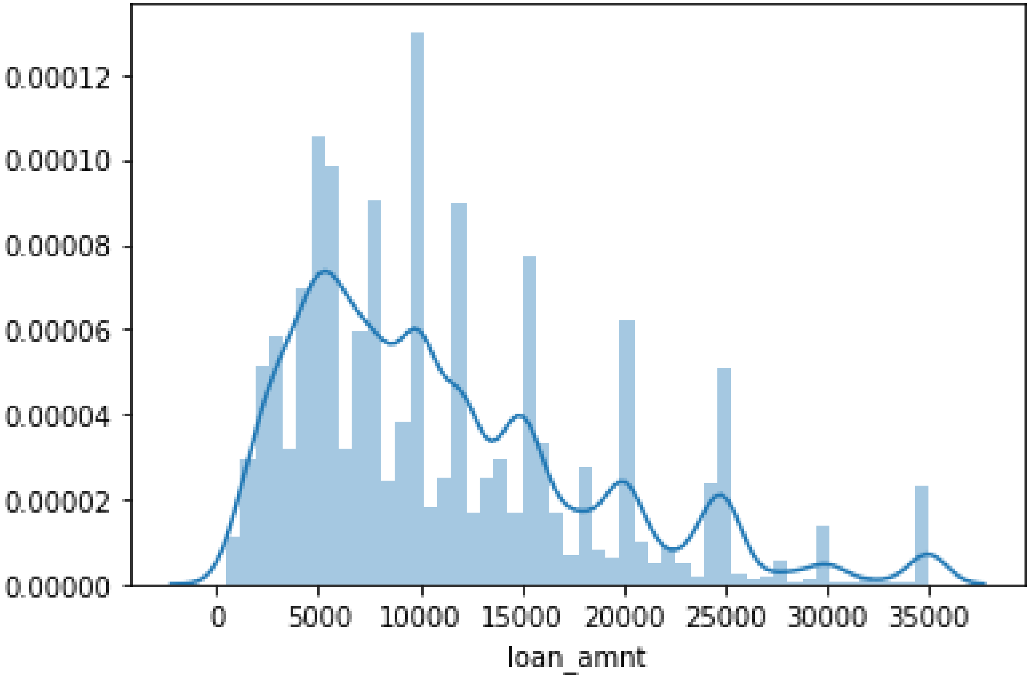


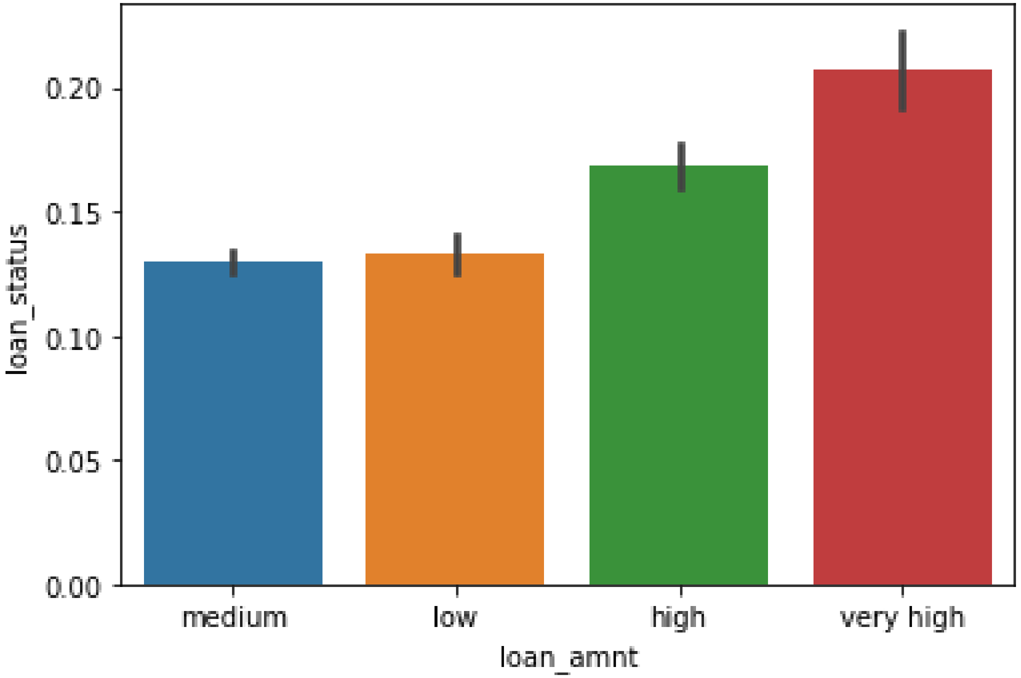
Inferences

Loans which were taken during the end of the year in months like from September to December are defaulted more , but there is not much variation across the months in the middle of the year.

Loan Amount vs Default Rate

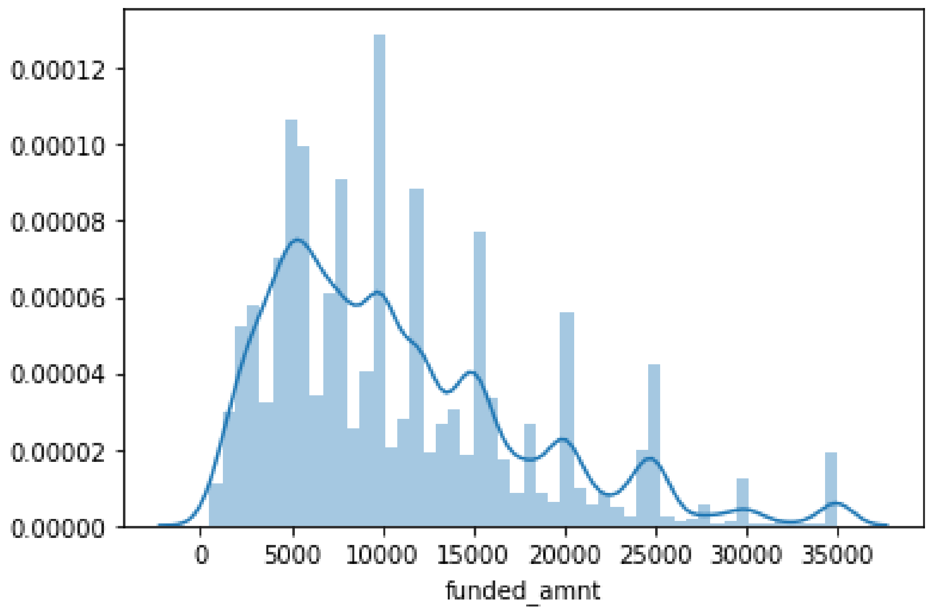
- Now we we will analyse how default rate varies for numerical variables.
- We have to use the binning concept for continuous numerical variables.
- Its clearly shown that: Greater the amount of loan, Greater the risk of default.
- First graph shows distribution of loan amount which can give idea about interval of binning.

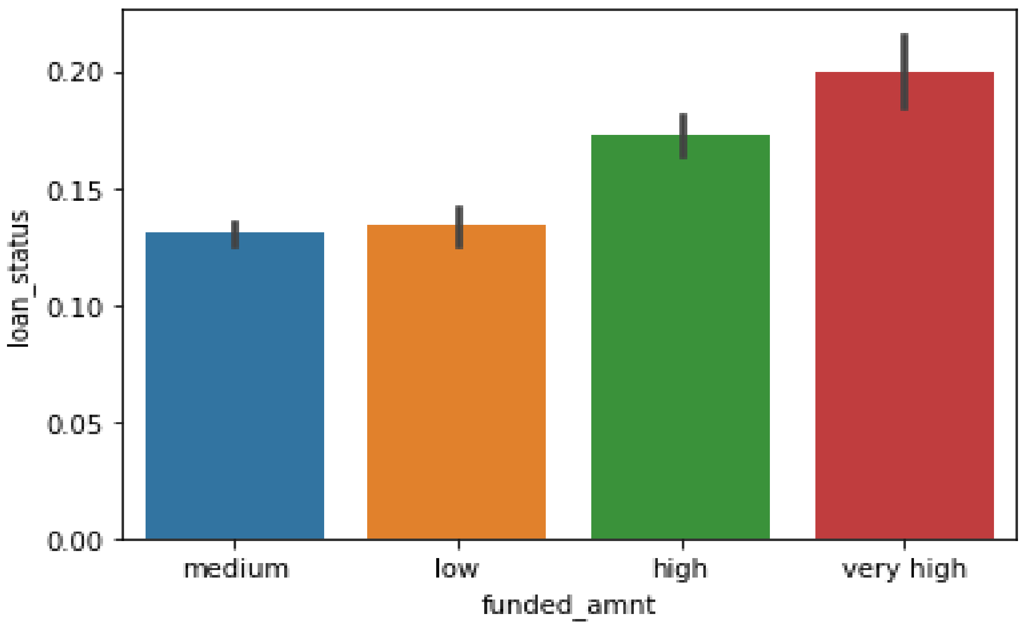




Funded Amount vs Default Rate

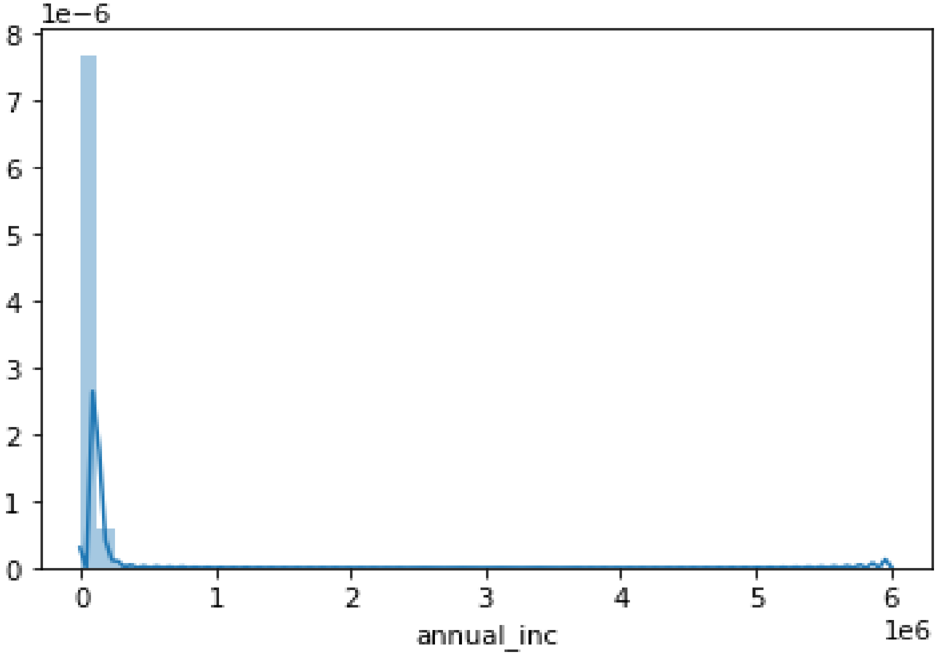
- Its clear from the graph below that more the funded amount more is the risk of default
- First graph shows distribution of funded amount which can give idea about interval of binning.

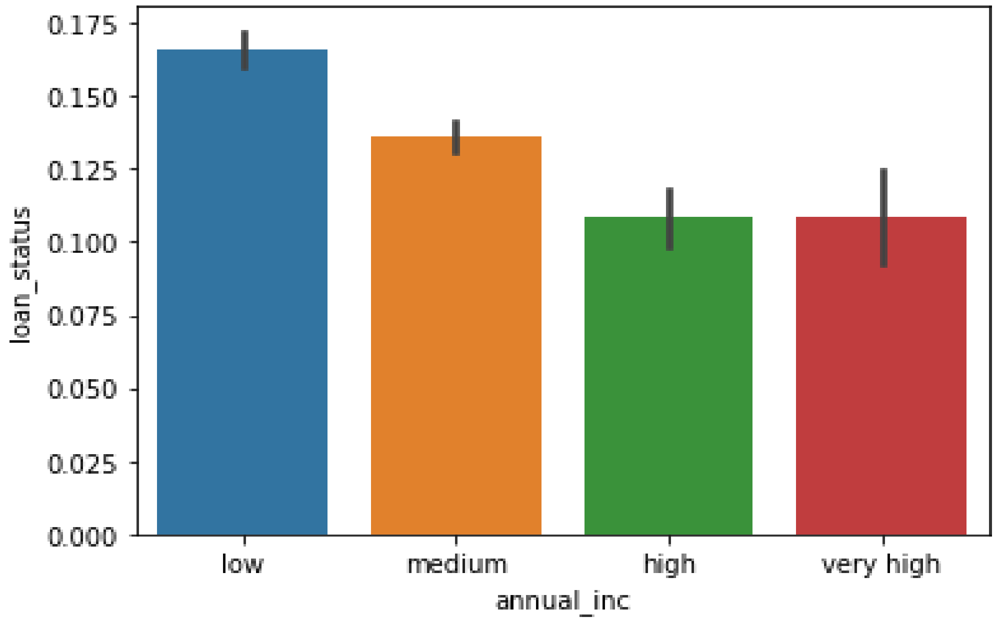




Annual Income vs Default Rate

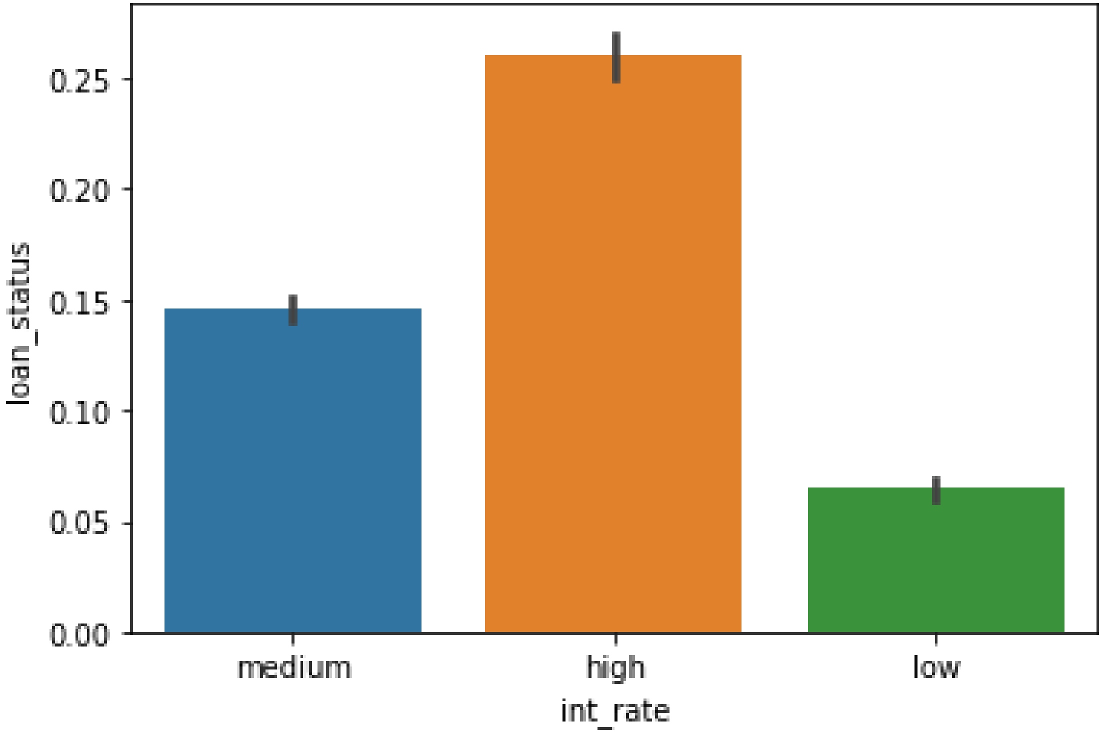
The first graph shows the distribution of annual income which help for creating bins
From second graph its clear lower the annual income higher is the risk of default.





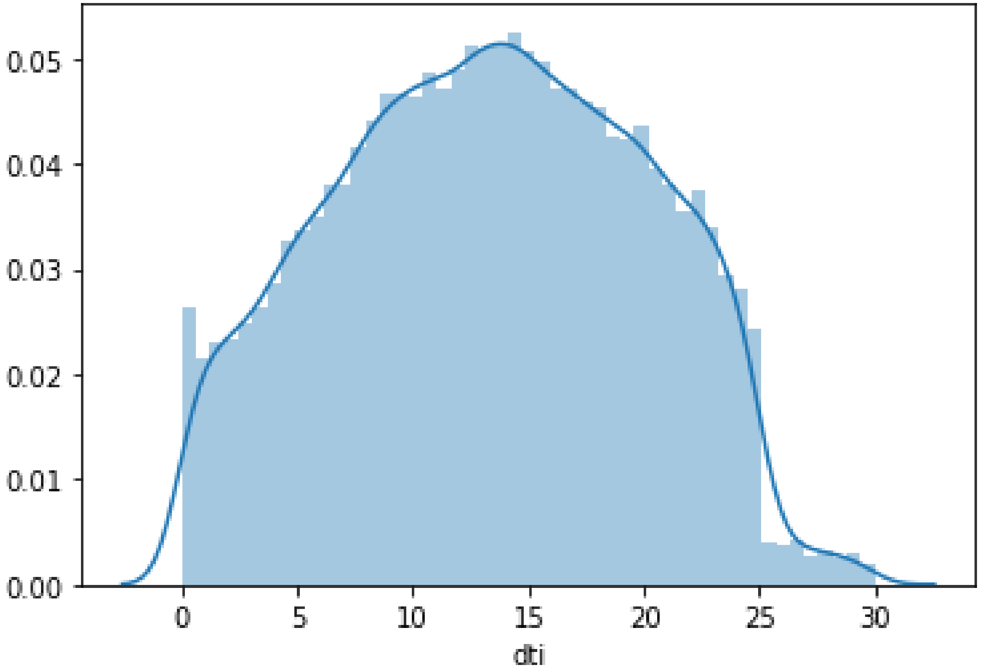
Interest Rate vs Default Rate

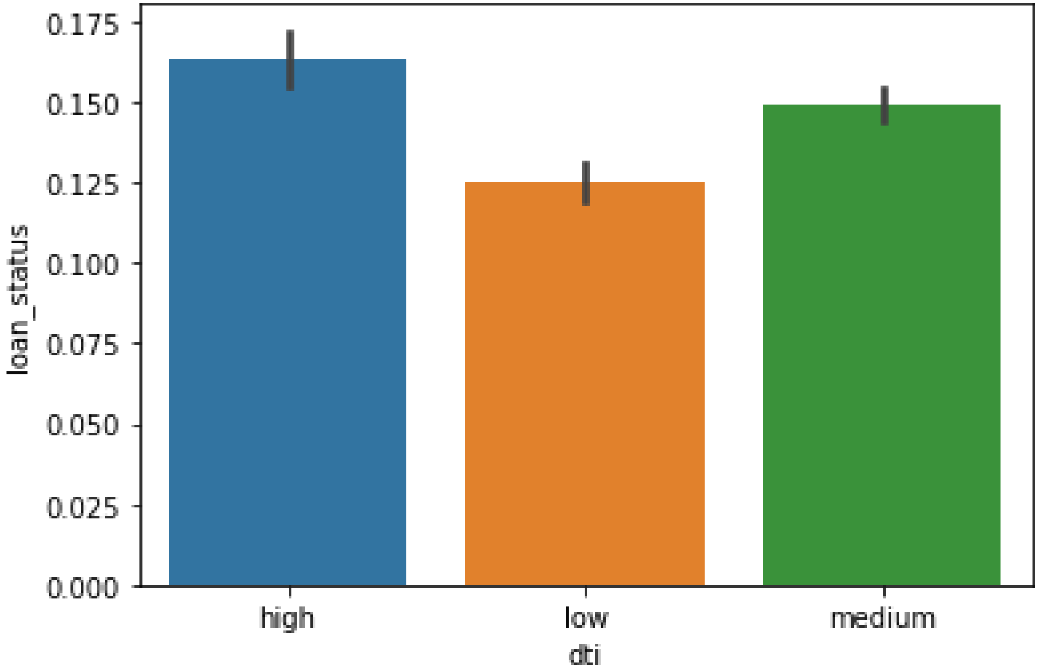
Inferences: Interest rate is directly proportional to default rate. i.e., Higher the interest rate higher is the risk of default.



Debt to Income (DTI) vs Default Rate

- First graph shows the distribution of dti which helps for binning.
- From second graph its clear that lower the dti ratio lower is the risk of default





Segmented Univariate Analysis

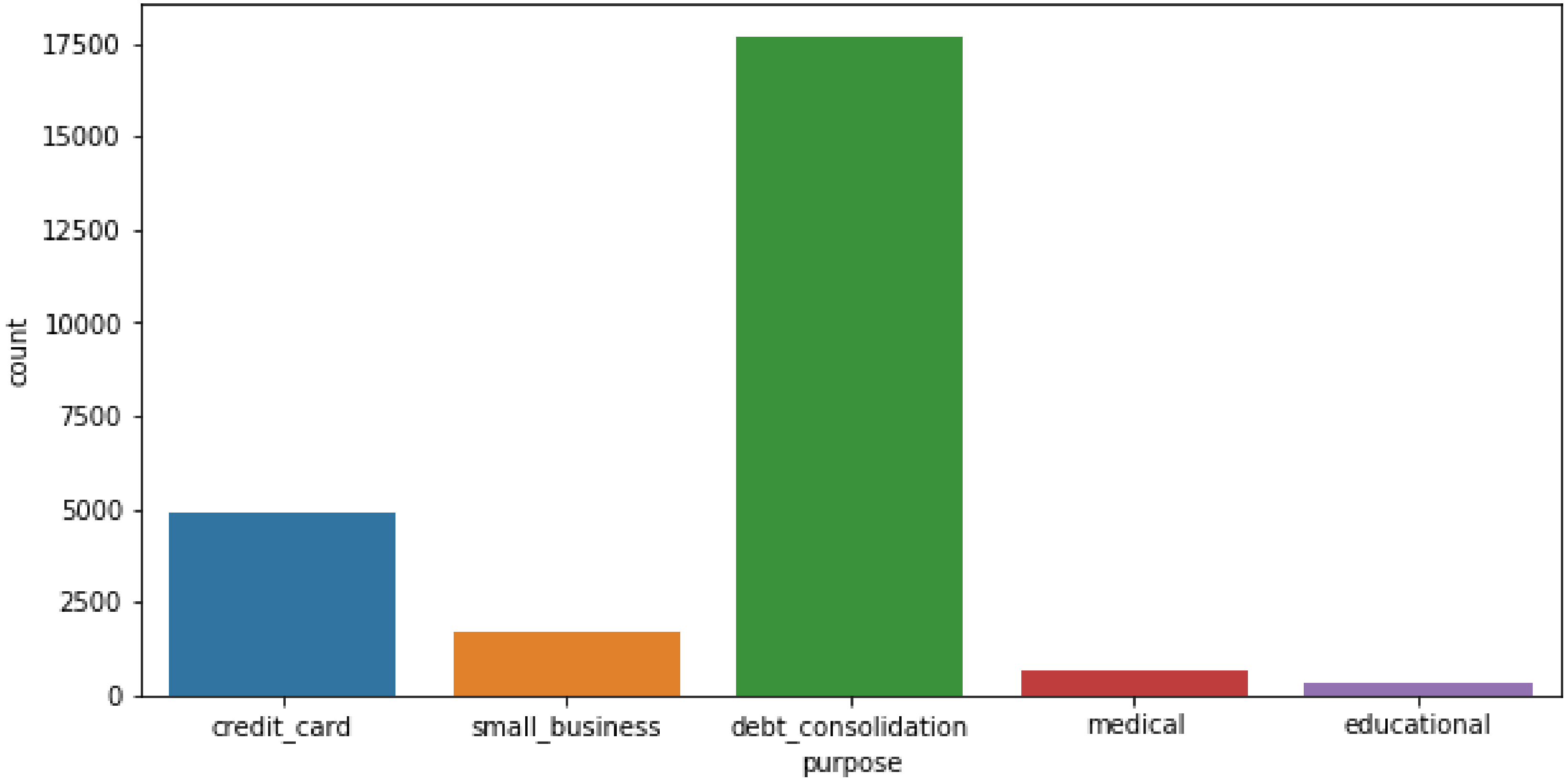
- We should compare the default rates across two variables and some of the important predictors are purpose of the loan, term ,interest rate, annual income, grade etc.

- Generally in the credit industry, one of the most important factors affecting default is the **purpose of the loan** - home loans perform differently than credit cards, credit cards are very different from debt consolidation loans etc.

- We will see how purpose of the loan is co related with other attributes like term , grade , sub_grade, loan_amount , etc.

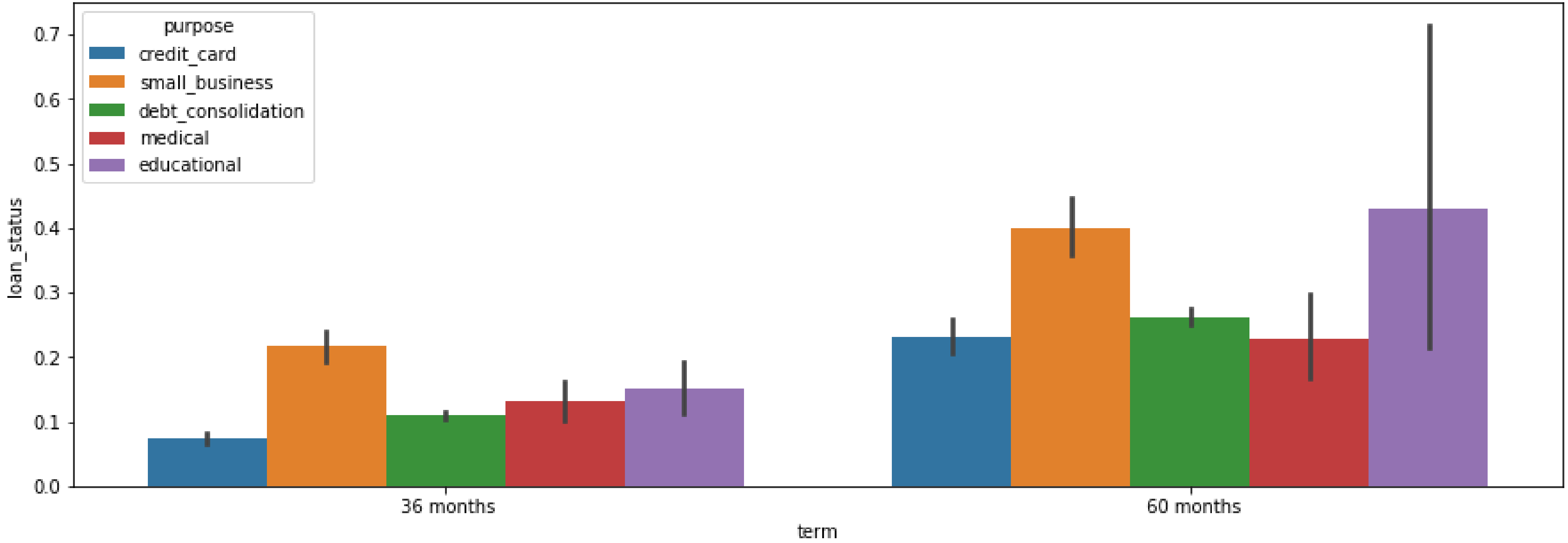
Plot that shows 5 Major Purpose of Loan

Purpose of loan vs count



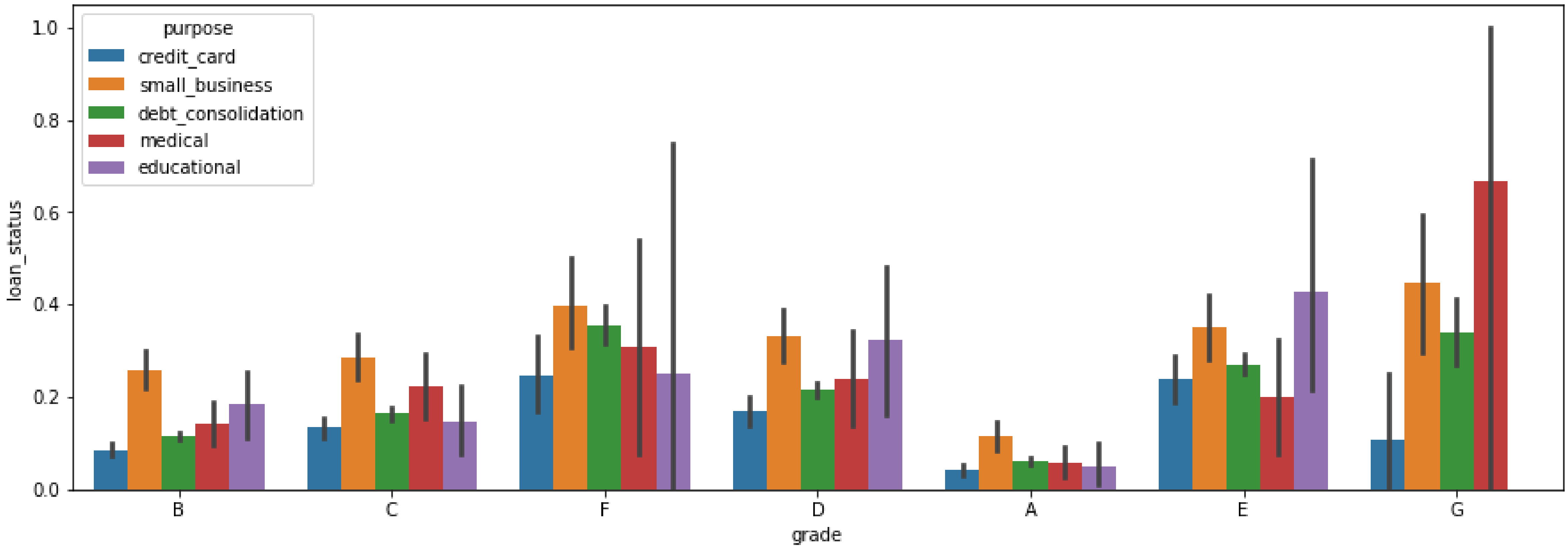
Bivariate Analysis

Default Rate with Purpose & Terms



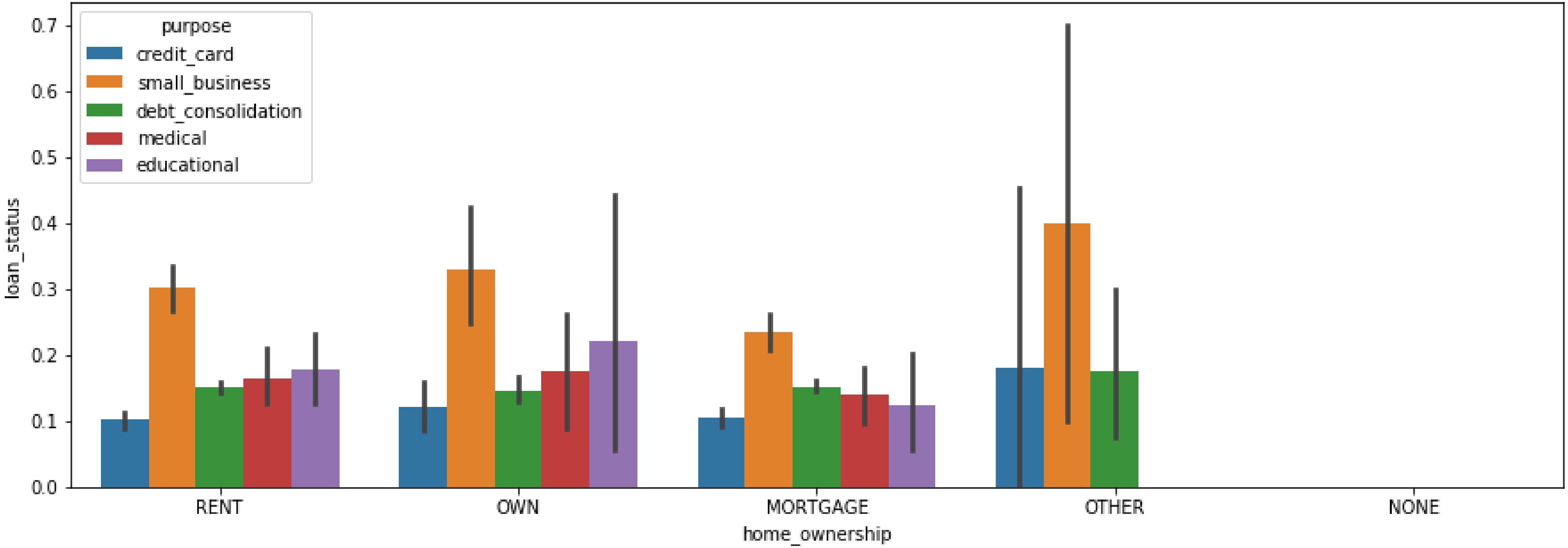
Inferences: We can find the insights between purpose of loan and the term. Clearly shows small business with 36 months term are more likely to be default .On the other hand for 60 months term educational loans are more likely to be defaulted.

Default Rate vs Grade With Purpose



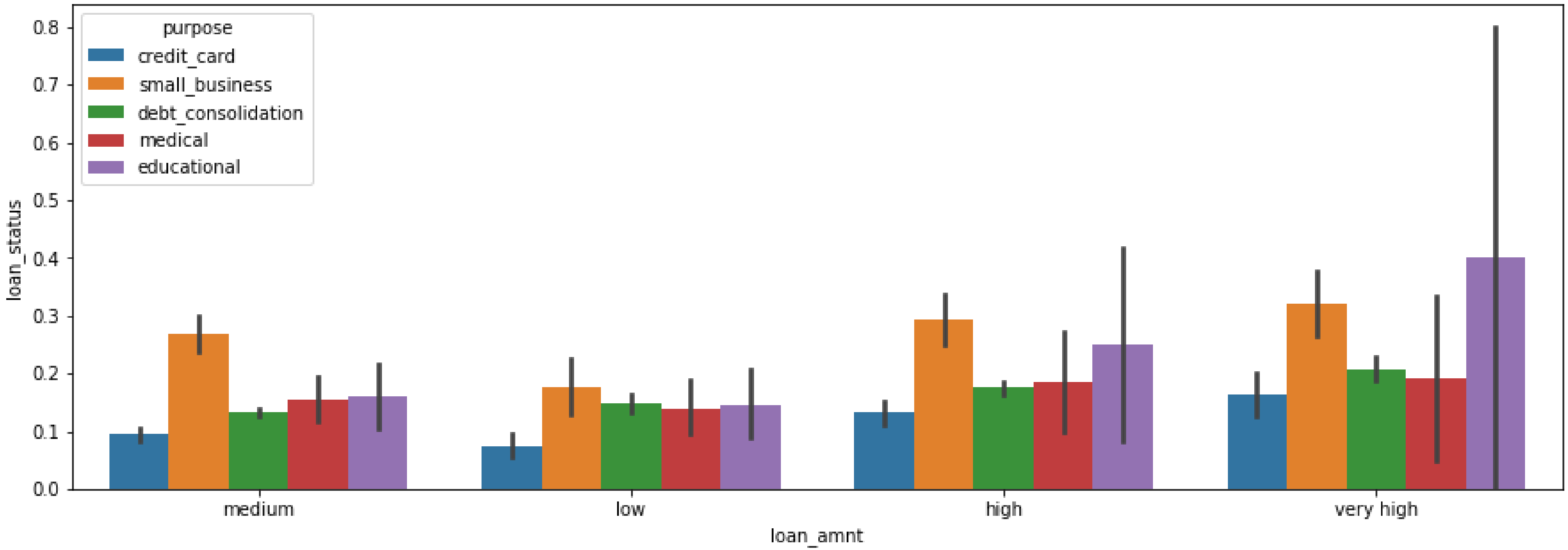
Inference: For A grades default risk is less for all types of the loan purposes which increase for B & which is less than C and so on.

Default Rate vs Home Ownership With Purpose



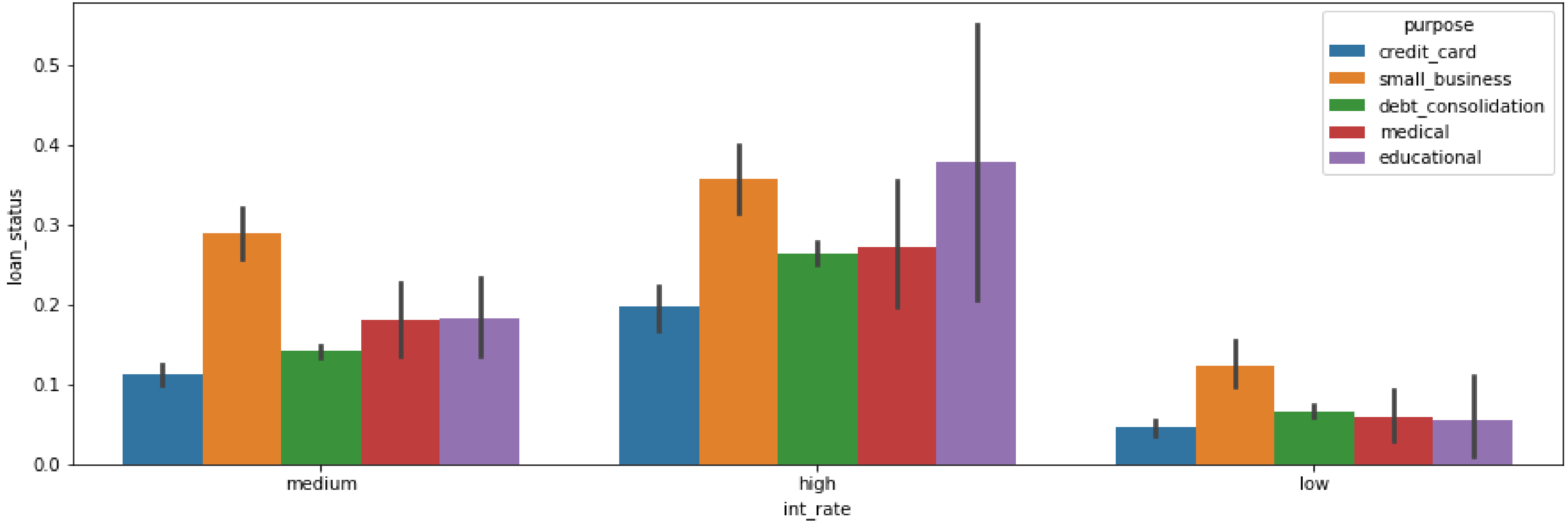
Inference : For "OWN " category of home ownership there are more chances of default.

Default Rate vs Loan Amount With Purpose



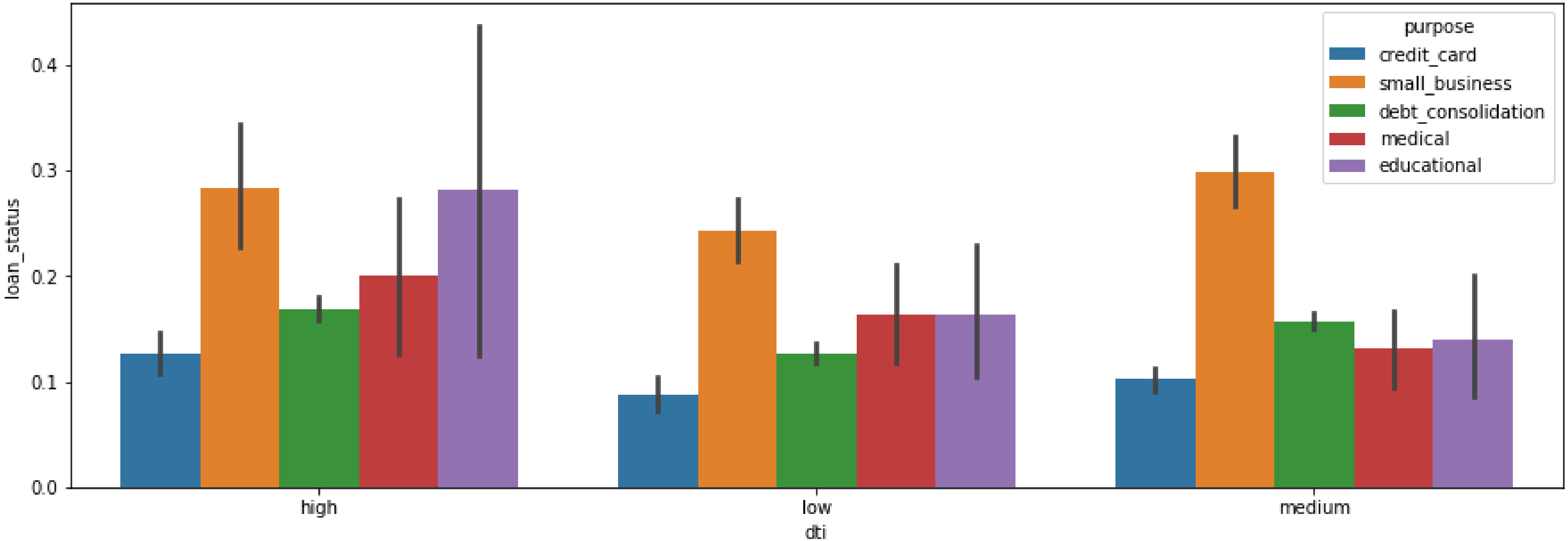
Inference : Higher the loan amount higher is the default rate. Education loans are more likely to default if loan_amount is high which is followed by the small_business.

Default Rate vs Interest Rate With Purpose



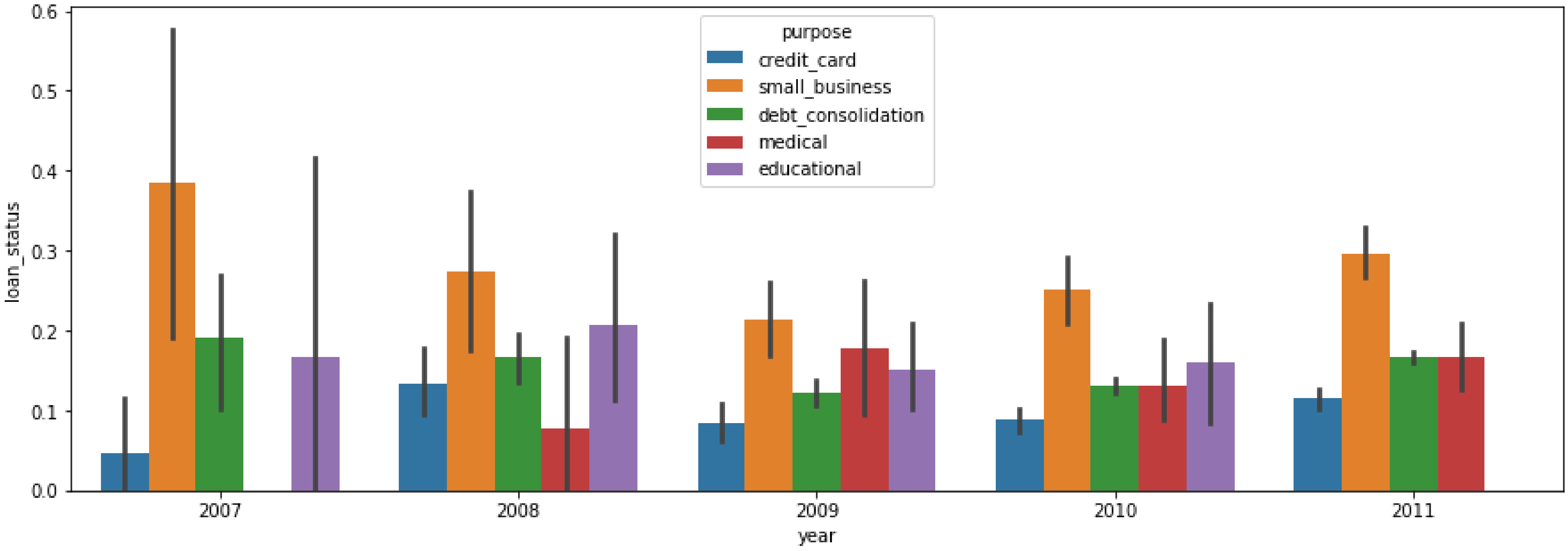
Inference : Higher the interest rate, higher chance of default rate. Also education loans are more likely to default if int_rate is high which is followed by the small_business and so on.

Default Rate vs Purpose with DTI



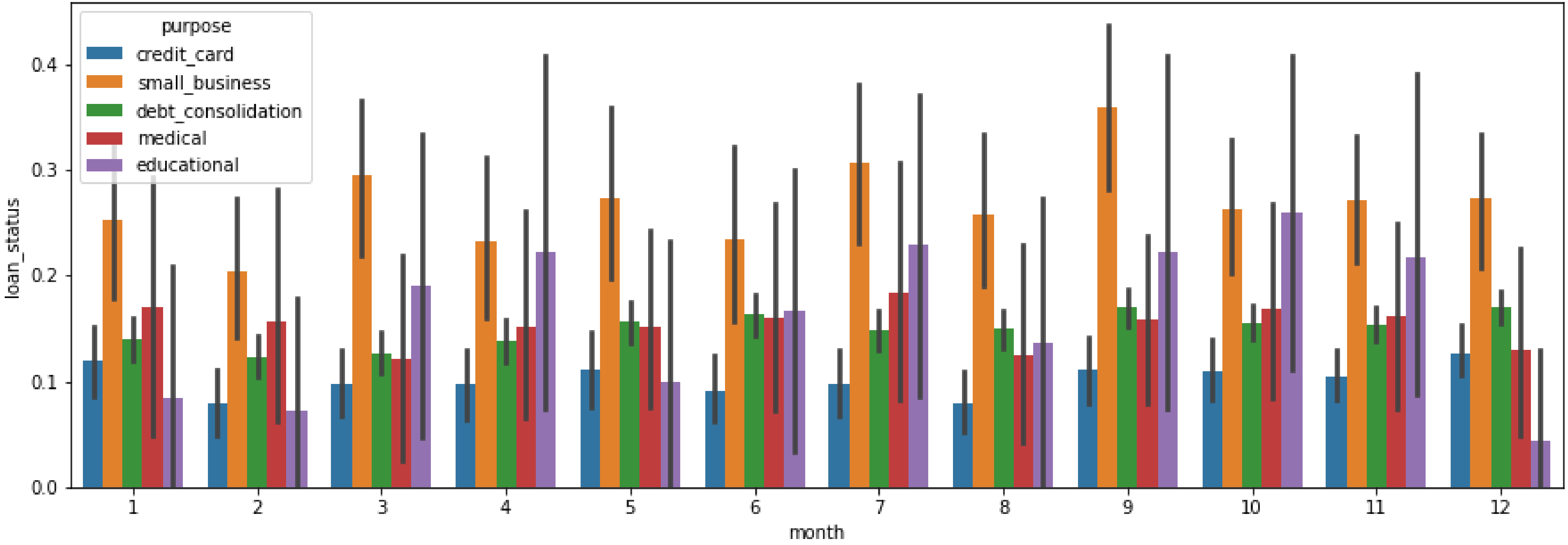
Inference : Higher the dti ratio, Higher the chance of default rate. So low dti ratio is safe for loan approval for the borrowers.

Default Rate vs Purpose with Year



Inference : From 2007 to 2011 , small_business are the main defaulters. Credit card loans seems to be the safest purpose.

Default Rate vs Purpose with Months



Inference : From Jan to December, small_business are the main defaulters. Also loans taken during the last quarter are more likely to default . Safest are the educational loans as per the above plot.

Conclusion

1. Loan characteristics such as amount of loan , interest rate , purpose of loan, etc are the major variables which are the predictors for the loan approval or rejection .

2. Some important observation are –
   - If Interest rate is high, the borrowers are more likely to default . Same pattern is seen for high loan amount , funded amount as well.
   - If dti ratio is large , default rate is high.
   - If annual income is low , default rate is high.
   - One interesting observation is seen for verification status of income source. Verified sources have high default rate .
   - The grades assigned to borrowers by Lending club can be used by investors to make a decision for approval or rejection of loans. ‘A ‘ grade borrowers are less likely to default as compared to B grade and C grade ones. So better grade have less default rate.
   - Same pattern is seen for sub grades as well .

To find the most Important predictors we have calculated the difference in the default rates for various attributes and the date is as follows :

- int_rate = 20%,
- home_ownership =19%,
- funded_amnt =7%,
- annual_inc = 6%,
- sub_grade =48%,
- grade = 29%,
- term= 16%,
- Dti = 3%,
- verification_status = 5%, l
- loan_amnt = 8%

So we can say that out of above variables -
int_rate, home_ownership, sub_grade, grade, term, loan_amnt can be taken as most important predictors as there Max and Min values
differ significantly.

So we can choose them as the **Most Important Predictors**.

Recommendations

1. Investors are recommended to keep in mind the following points while making a decision for loan approval or rejection :

   - Check the grades and sub grades assigned by the Lending club to the loan applicant . Applicants with lower grades (A<B<C….) are safe to be given loan as their default rate is low.
   - Applicants with high annual income are less likely to default.
   - If the interest rate is high there are chance that loan will be defaulted(Same pattern is seen in case of loan amount , funded amount as well) . So If we see combination of different factors , for example if grades are good , set less but adequate interest rate.
   - Loan approved for those with less dti(debt to income ratio) are much more safe as compared to high dti ratio.
   - If loans are given for small business then keep interest rate on higher side , as in past loans taken for small business are defaulted most.