Lending Club Case Study

Manish Tyagi & Prashanthi Chapala

Agenda

Background & objective

Approach & Methodology

Data Insights

Recommendations

1. Background & Objective

Background

- ✓ Lending club is an online platorm that brings borrowers and lenders(investors) together.
- ✓ Borrowers can apply for various kinds of loans e.g. Personal, business, auto refinancing,etc.
- ✓ Investors can decide whether to finance a loan or not.

Objective

- ✓ Identify the driving factors (or driver variables) behind the loan default.
- ✓ Using these driving factors, the lending club can idetify the 'risky' loan takers who have the higher chance of defaulting.

2. Approach & Methodology

Approach

- ✓ Initial understanding of the data & cleansing to rectify any data issues.
- ✓ Classify the variables based on business knowledge into relevant and non-relevant.
- ✓ Univariate & segmented univariate analysis of default % across various variables.
- ✓ Bivariate analysis of default % across various continous & categorical variables.

Methodology

- ✓ The variables can be classified broadly into
    1. Related to the applicant (demographic variables such as age, occupation, employment details etc.),
    2. Loan characteristics (amount of loan, interest rate, purpose of loan etc.) and
    3. Customer behavior variables (those which are generated after the loan is approved such as delinquent 2 years, revolving balance, next payment date etc.).
- ✓ Customer behavior variables are not available at the time of loan application, and thus they cannot be used as predictors for credit approval.
- ✓ Only the records where loan status is either 'Charged Off' or 'Fully Paid' will be considered for analysis.
- ✓ Funded Amount Investors is used for analysis as its the amount sponsored by the investors.

Methodology(contd..)

- Few continous variables relevant for analysis were converted to categorical variables by 'binning'.
- Below is a summary of data buckets that were used in the 'Binning' of variables :

Funded Amount Investor

Less than 5000 is 'Low'
Greater than 5000 and less than 15000 is 'Medium'
Greater than 15000 and less than 25000 is 'High'
Greater than 25000 is 'Very High'

Interest Rate

Less than 10 is 'Low'
Greater than 10 and less than or equal to 15 is 'Medium'
Greater than 15 is 'High'

DTI(debt to income ratio)

Less than 10 is 'Low'
Greater than 10 and less than or equal to 20 is 'Medium'
Greater than 20 is 'High'

Annual Income

Less than or equal to 50000 is 'Low'
Greater than 50000 and less than equal to 100000 is 'Medium'
Greater than 100000 and less than or equal to 150000 is 'High'
Greater than 150000 is 'Very High'

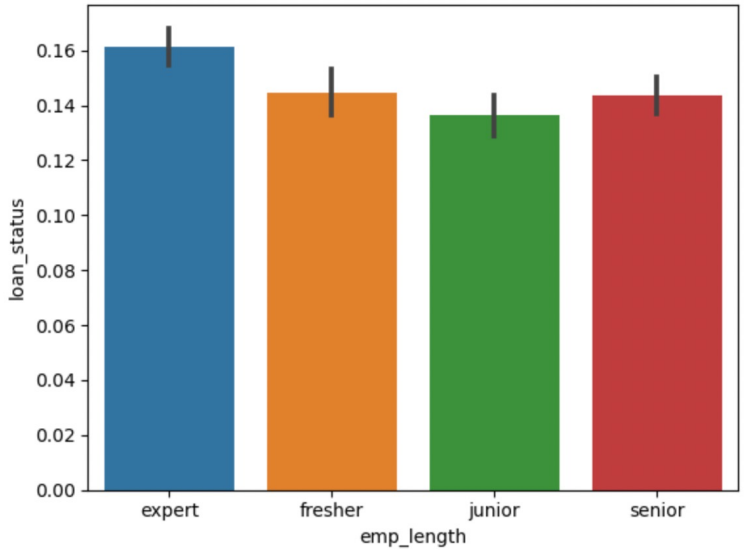
Employment Length

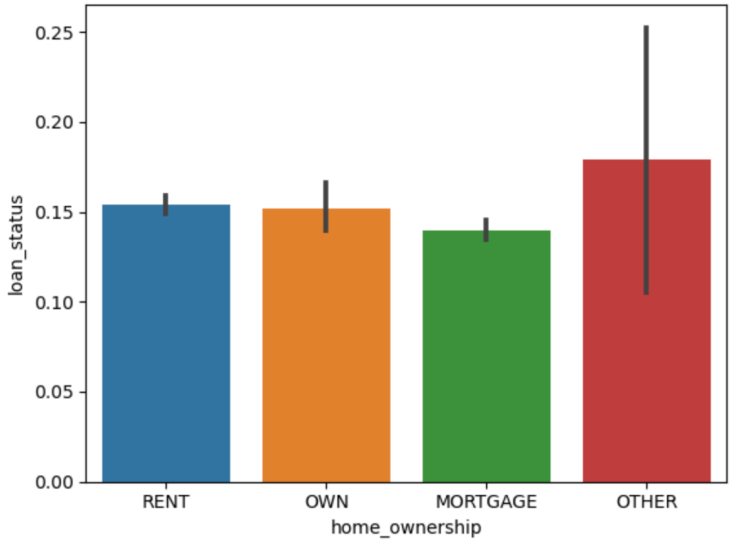
Less than or equal to 1 is 'fresher'
Greater than 1 and less than equal to 3 is 'junior'
Greater than 3 and less than or equal to 7 is 'senior'
Greater than 7 is 'expert'

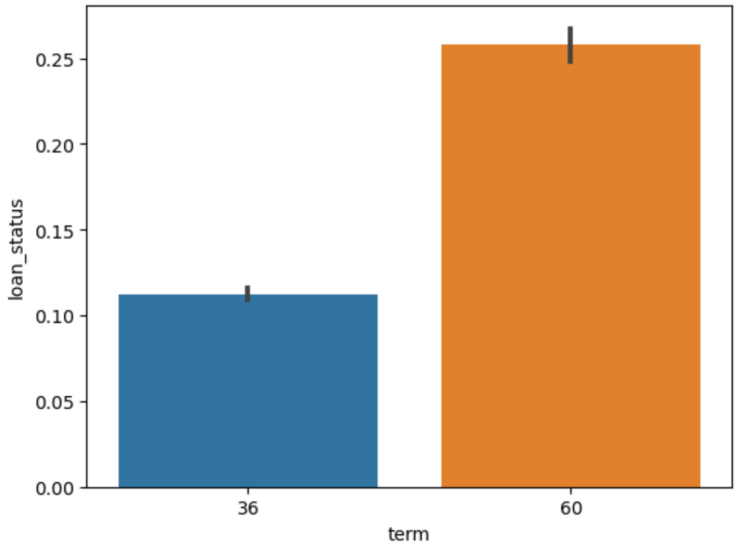
2. Data Insights

Evaluation of variables

- ✓ Variables ‘employee length’,’home ownership’ cant be good indicators for default as across all values for these variables, there is no considerable difference in default rate.
- ✓ Variable loan term ‘term’ is a strong indicator for default% as difference between 36 months and 60 months is considerable.

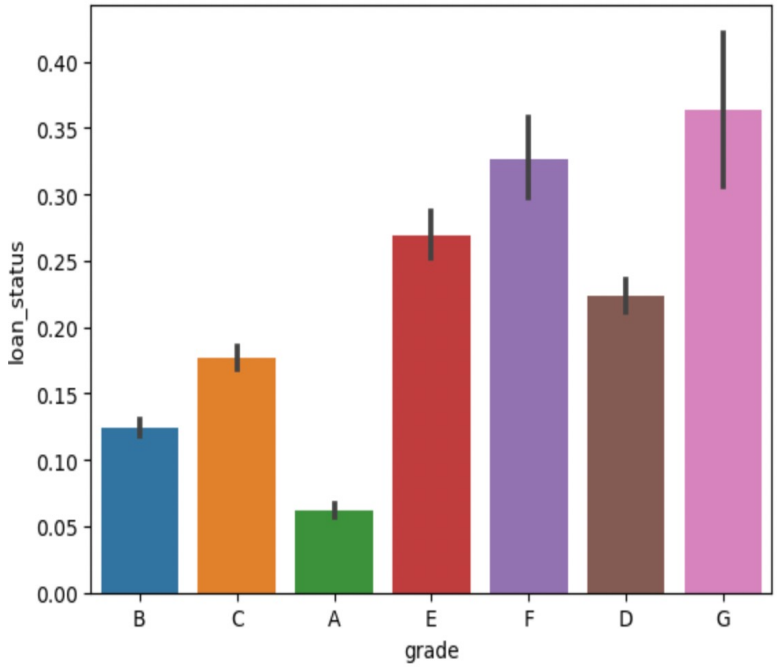


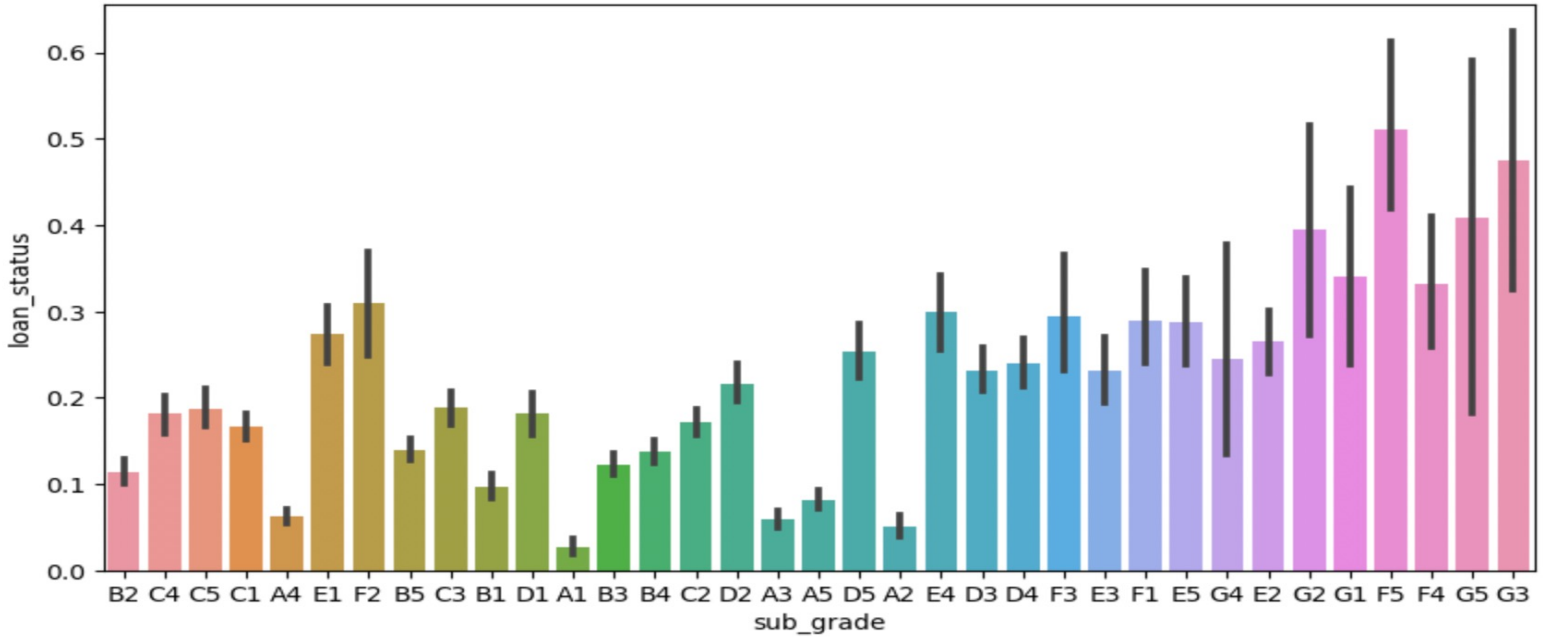




Evaluation of variables (contd..)

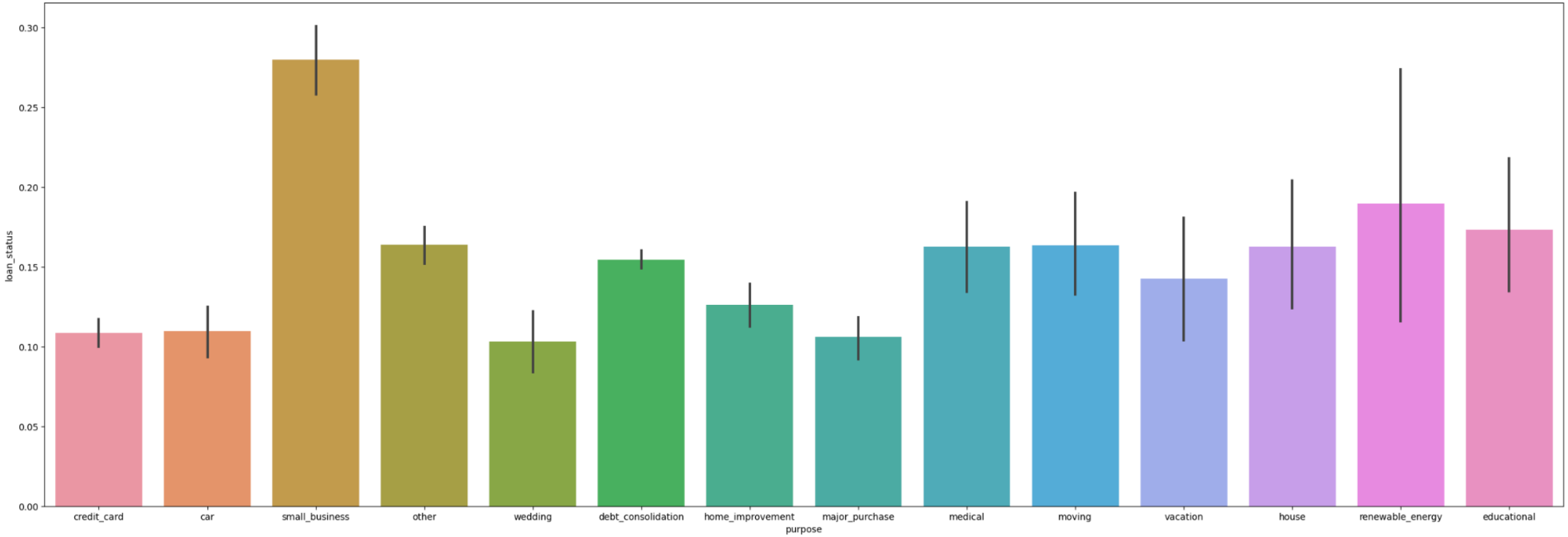
- ✓ Variables ‘grade’ and ‘sub_grade’ show expected results, i.e. low grades have higher loan default rate .
- ✓ These variables can be considered as an indicator for loan default and will be considered for further analysis.





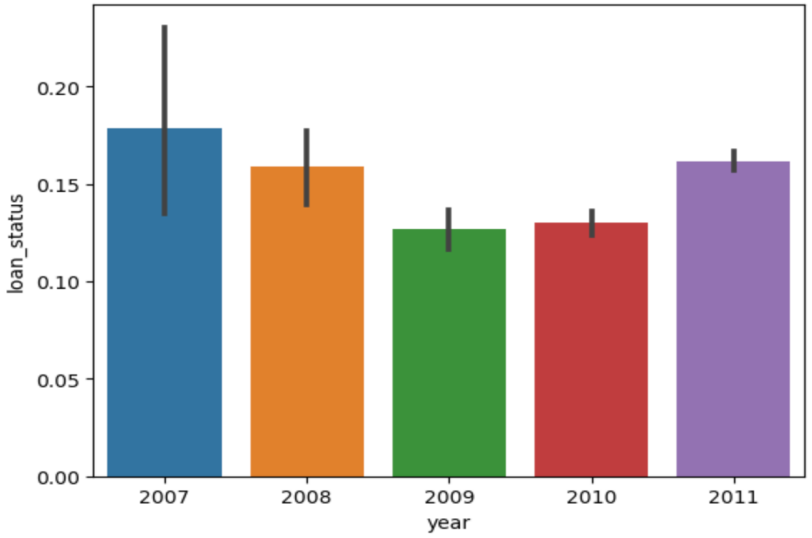
Evaluation of variables (contd..)

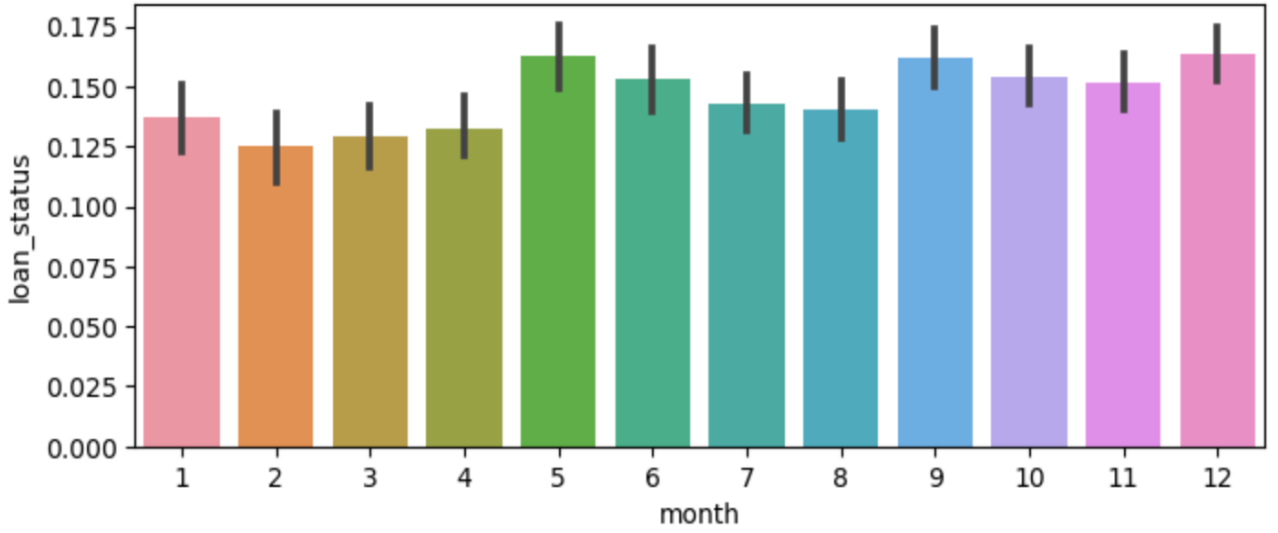
- ✓ Variables 'purpose' shows that small business, renewable energy, educational and house the have highest default %.
- ✓ This variable can a good indicator for default and will be considered for further segmented & bivariate analysis with other important variables.



Evaluation of variables (contd..)

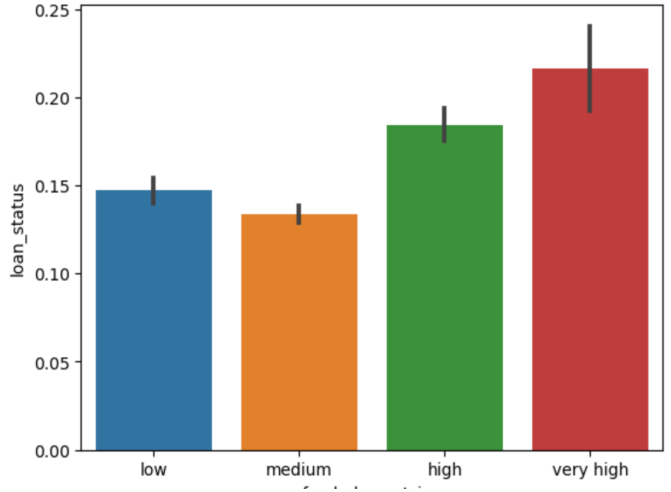
- ✓ Variables ‘year’ shows that default % decreased every year from 2007 till 2010, but again jumped up in 2011. The pattern is not strong enough for this variable to be considered as an indicator for default.
- ✓ Variable ‘month’ shows that default rate is highest in Dec, Sep and May.
- ✓ Since the difference between the default rate for various months is not considerable, this cant be a strong independent indicator for default but can still be considered for analysis along with other relevant variables.

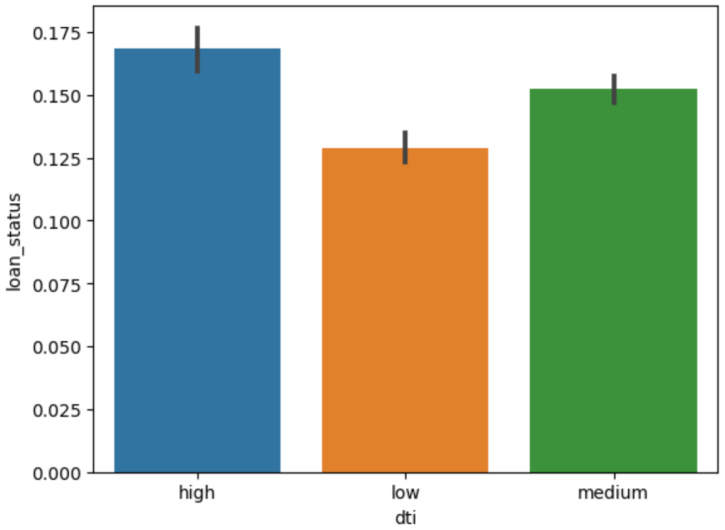


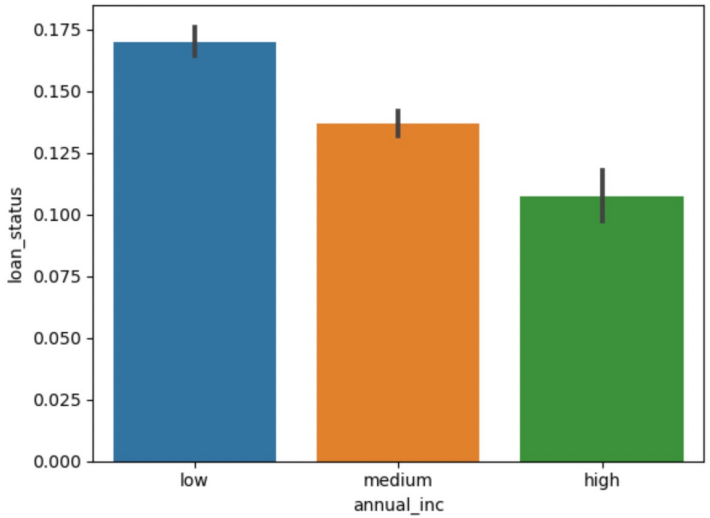


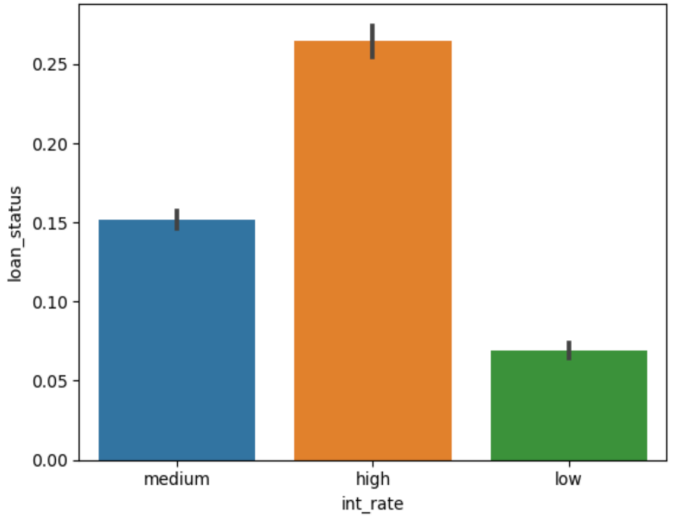
Evaluation of variables (contd..)

- ✓ Variables ‘funded_amnt_inv’, ’int_rate’ and ‘annual_inc’ show that loan default % has considerable difference between value categories. Hence, these are good indicators for loan default and could be considered for further analysis along with other variables.
- ✓ Variable ‘dti’ shows that value category ‘high’ has the highest default rate. But the difference between various category values is not very high, but still it can be considered for further segmented and bivariate analysis along with other variables.







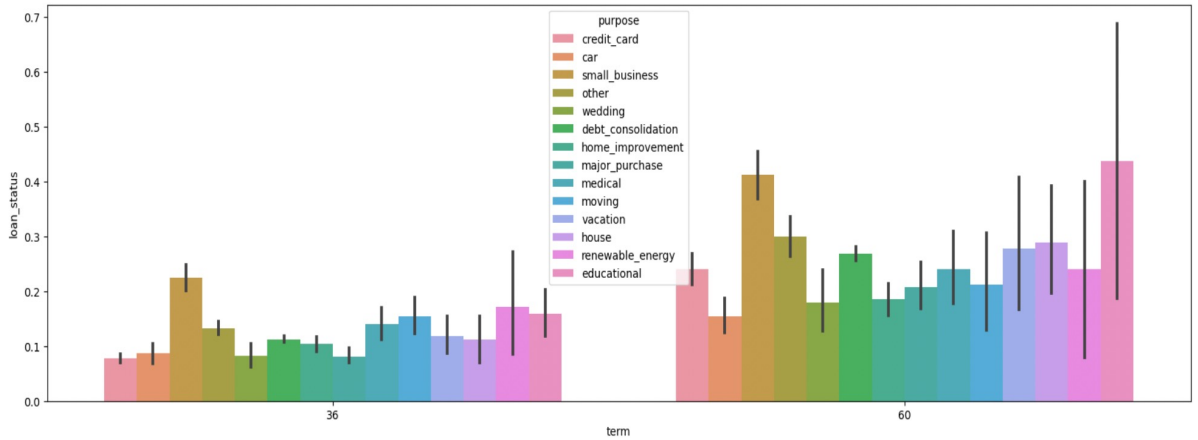


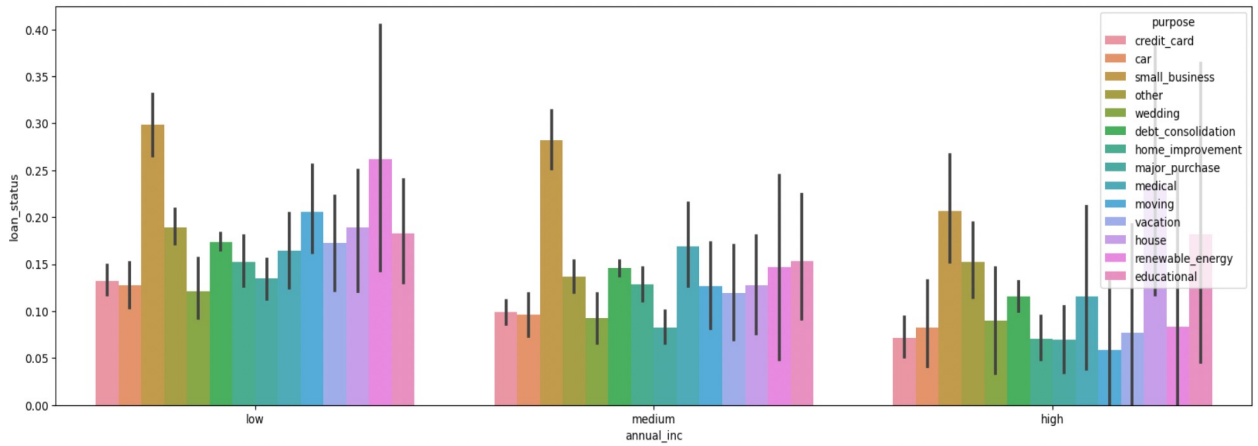
Summary of Variables

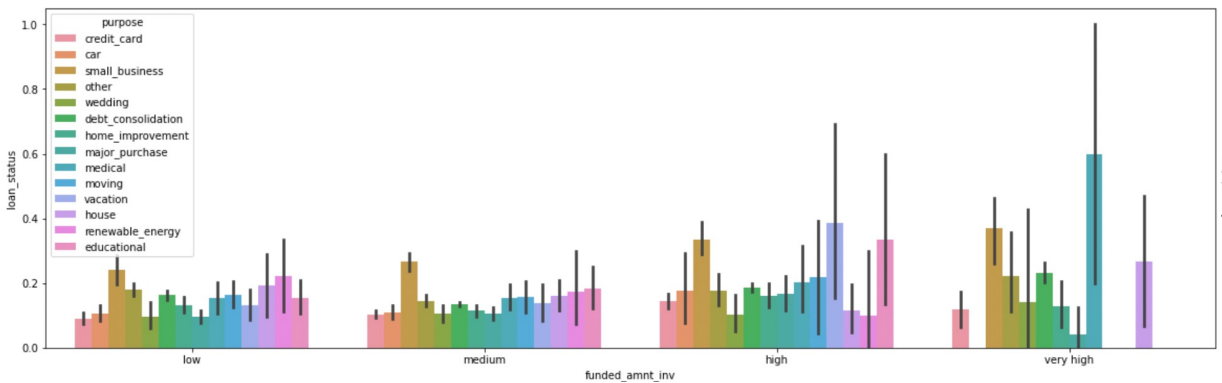
- ✓ Variables with higher difference % between highest and lowest default rate are considered strong indicators of loan default.
- ✓ Variables sub grade, grade, int_rate, purpose, term, funded_amnt_inv and annual_inc are the strongest indicators of loan default.
- ✓ Variables with lower difference % between highest and lowest default rate can be used in further segmented analysis of default rate.

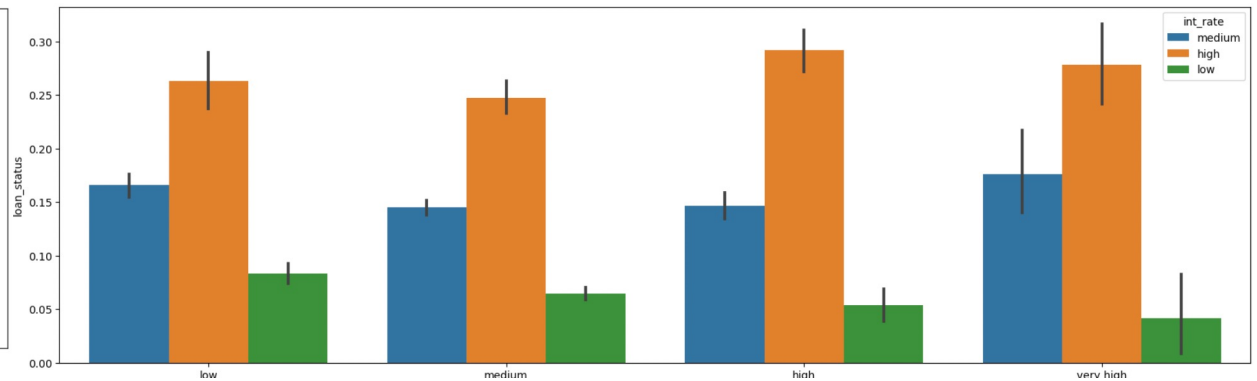
Variable	Approx. Difference between highest & lowest default rate	Remarks
Grade	30%	% difference between highest & lowest default rate is considerable. These variables could be a good indicator of loan default. Slicing & dicing of these variables along other variables could give useful insights about loan default rate.
Sub Grade	30%	
Interest Rate	20%	
Purpose	17%	
Term	15%	
Funded Amount by Investors	10%	
Annual Income	7%	
Month	5%	
Verification Status	4%	% difference between highest & lowest default rate is bot considerable. But these variables can be analysed along with other variables to get useful insights about loan default.
DTI	4%	
Home Owenership	3%	
Year	3%	
Employee length	2%	

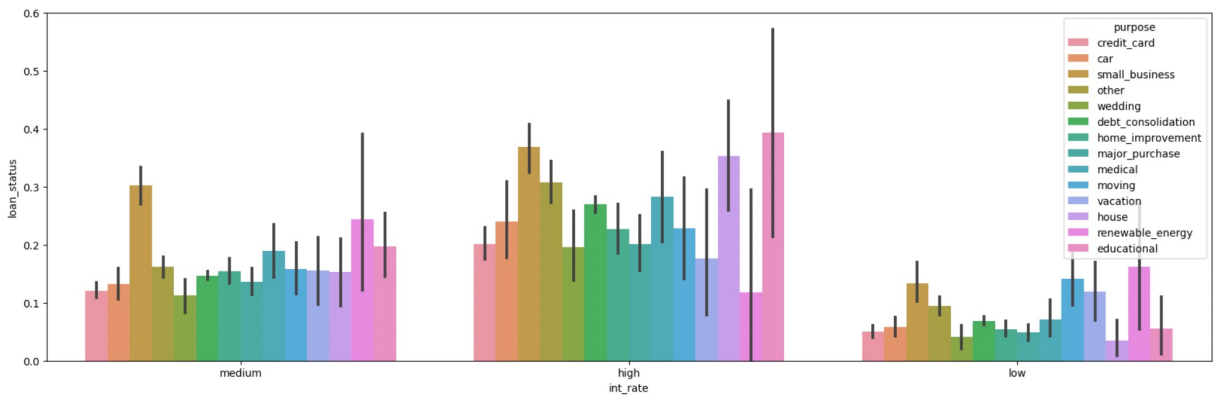
Data Insights

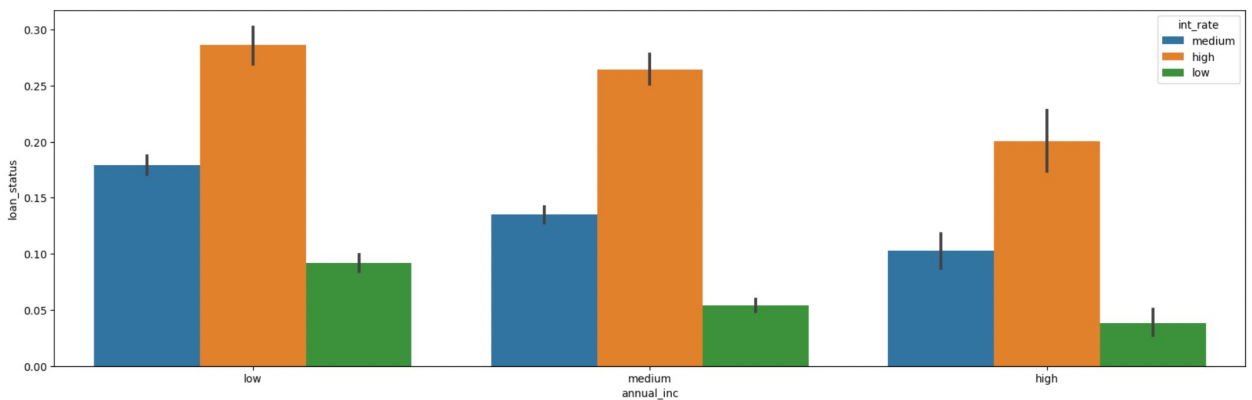












4. Conclusions & Recommendations

What does the data tell us about loan defaults?

✓Of all loan applicants, 14% default.
✓Loans with a term duration of 60 months have the highest default rate.
✓Lower loan grades D,E,F & G and lower sub-grades like F4,F5,G3 & G5 have the highest default rate.
✓When loan purpose is 'Small business', 'Renewable energy' & 'Educational', the default rate is the highest.
✓When the loans are either Verified or Funded amount by investor is very high(>25000) or the Interest Rate is high(>15%) or the DTI is high(>20) or the Annual Income is low(<50000) and the employment length of the applicant is expert(>7 years), the default rate is the highest in each categories.
✓When the loan purpose is 'Small business' , for every bucket in Term, Funded amount by investors, DTI, and Annual Income, the loan default rate is the highest.
✓When Interest Rate is high(>15%) and term is 60 months , the loan default rate is highest at 35%.
✓When the funded amount by investors is very high and interest rate high , the loan default rate is close to 30%
✓Low annual income and high interest rates have an approx. 30% loan default rate.
✓In every month and in every annual income group, high interest rates have the highest loan default rates.
✓Across all loan verifications status and DTI buckets , high interest rate has the highest loan default rates.

Final Recommendations

Following driver variables can best help predict a loan default:

1. Loan Term,
2. Loan Purpose,
3. Interest Rate,
4. Funded Amount by Investors,
5. Loan Grade,
6. Annual Income
7. DTI

Other variables like verification status, Month, Employee Length can be helpful in further slicing and dicing the data to get useful insights about loan default rate.