Lending Club: Case study

SUBMISSION

Name:

1. Harsha Vardhan Reddy
2. Vignesh Ramaya

Lending Club

What is Lending Club?

Lending Club is a marketplace for personal loans that matches borrowers who are seeking a loan with investors looking to lend money and make a return.

How Lending Club works?

1. Customers interested in a loan complete a simple application at LendingClub.com.

2. Lending Club evaluates each borrower's credit score using past historical data (and their data science process!) and assigns an interest rate to the borrower.

3. Qualified applicants receive loan offers in just minutes and can evaluate their options with no impact to their credit score.

4. Investors select the loans they want to invest in based on their own risk tolerance, investment portfolio goals, and time horizon.

Problem Statement

Whenever a loan application is filed then there are 2 types of risks associated

1. If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company

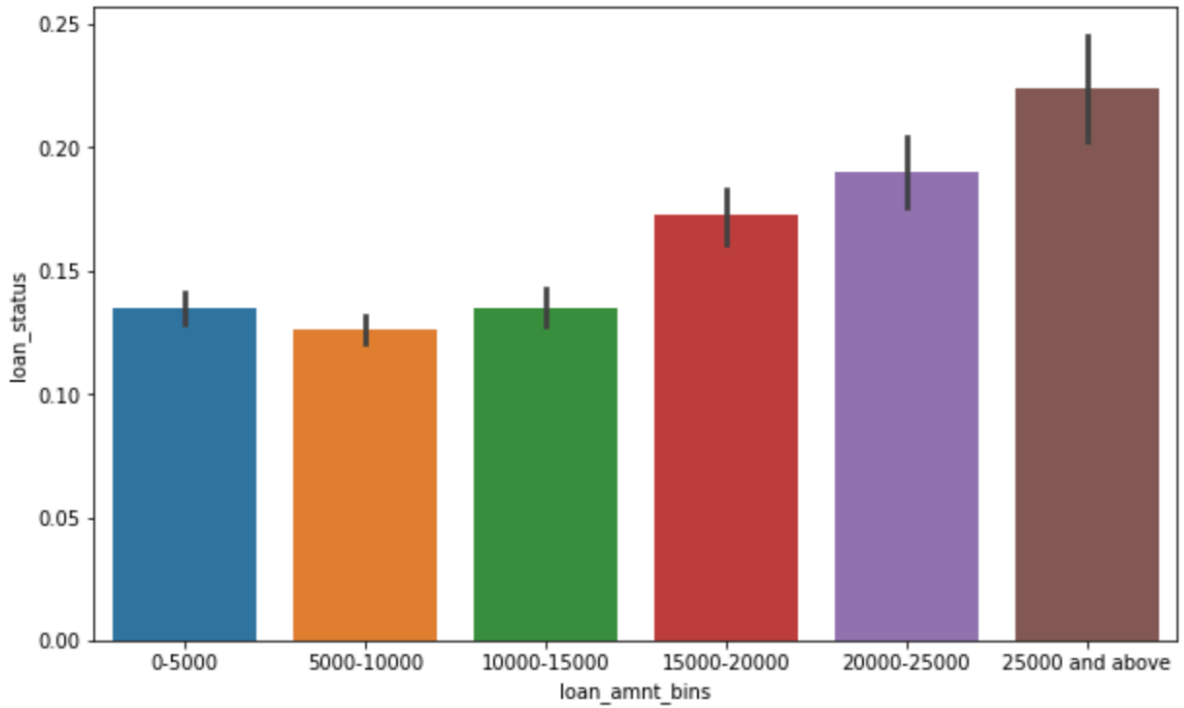
2. If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company

For lending club, the second factor, that is lending loans to ‘risky’ applicants is the largest source of financial loss (called credit loss).
Hence we have to identify the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default.

EDA Analysis

Univariate analysis

1. Loan Amount:



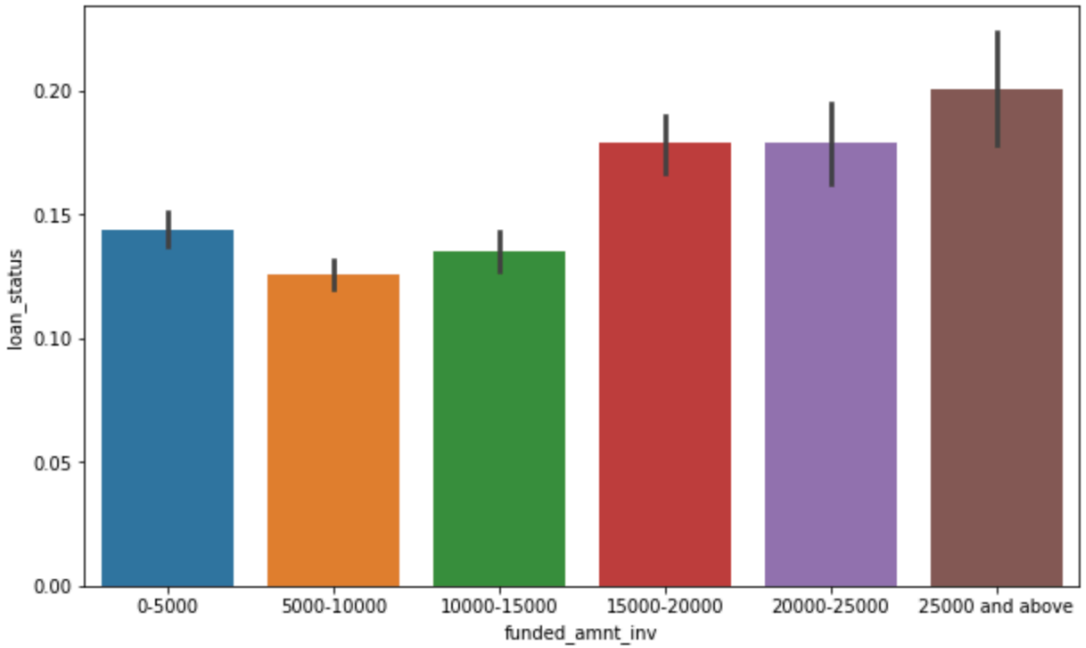
Observation

You can notice apart from first bin, as the loan amount increases the default rate increases as well

EDA Analysis

Univariate analysis

2.Funded amount Invested :



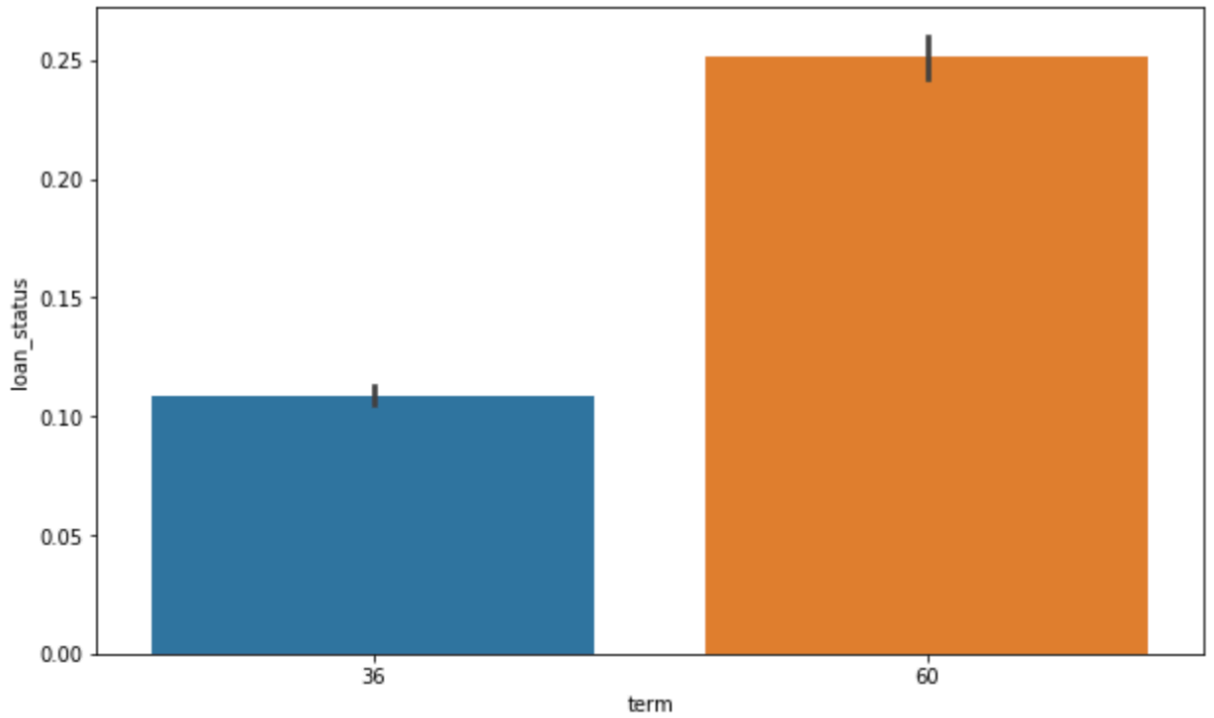
Observation

follows the same pattern as the loan amount, that is, as the funded amount increases so does the default rate
It simply meams that higher loan amounts are likely to get charged of compared to lower loan amounts

EDA Analysis

Univariate analysis

3. Term:



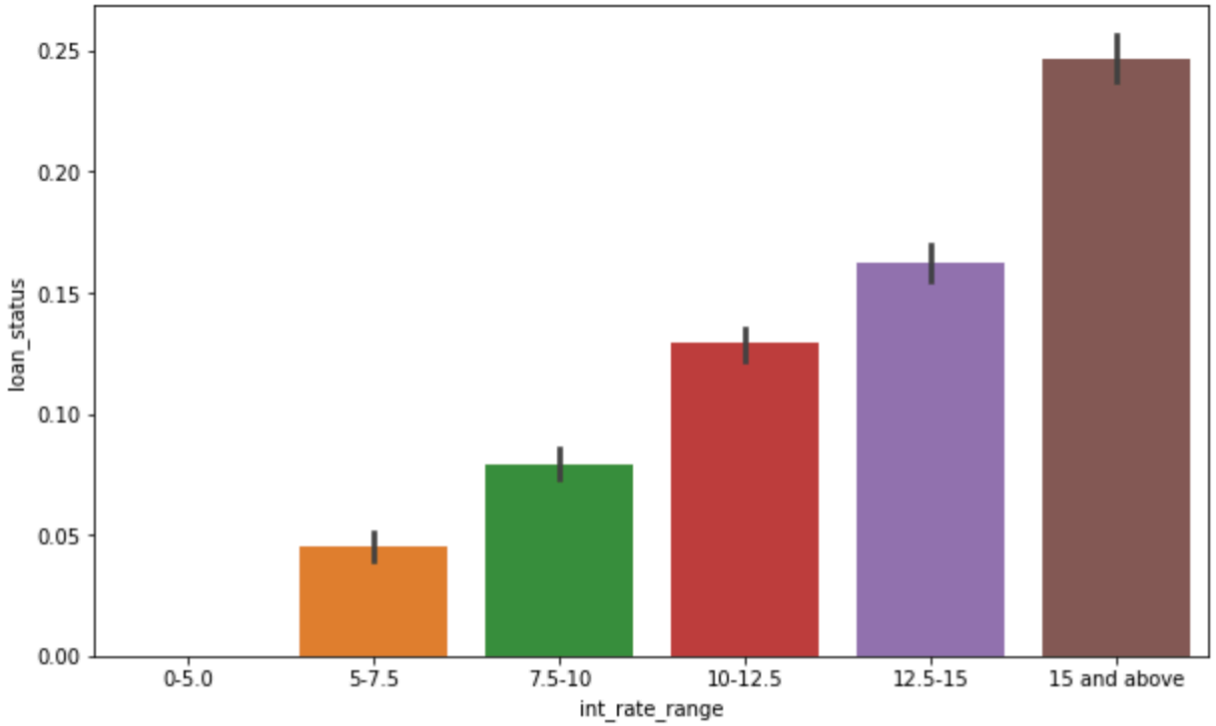
Observation

Longer the loan term higher is the default rate

EDA Analysis

Univariate analysis

4. Interest rate:



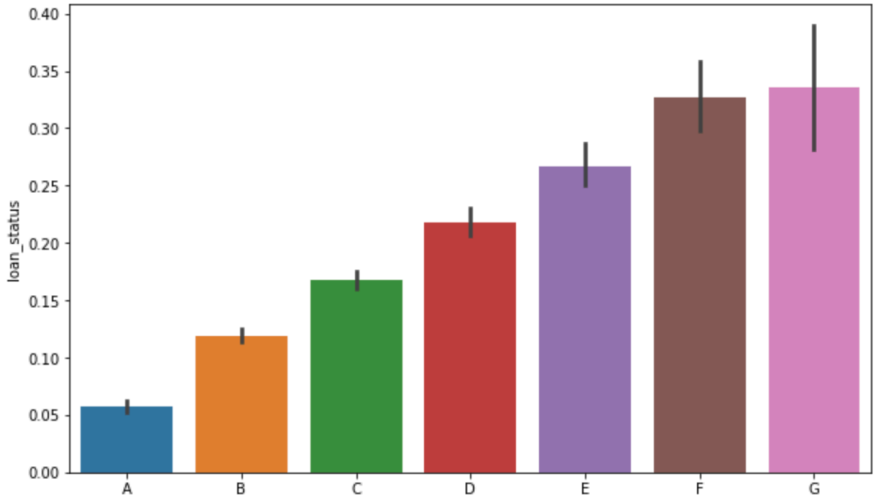
Observation

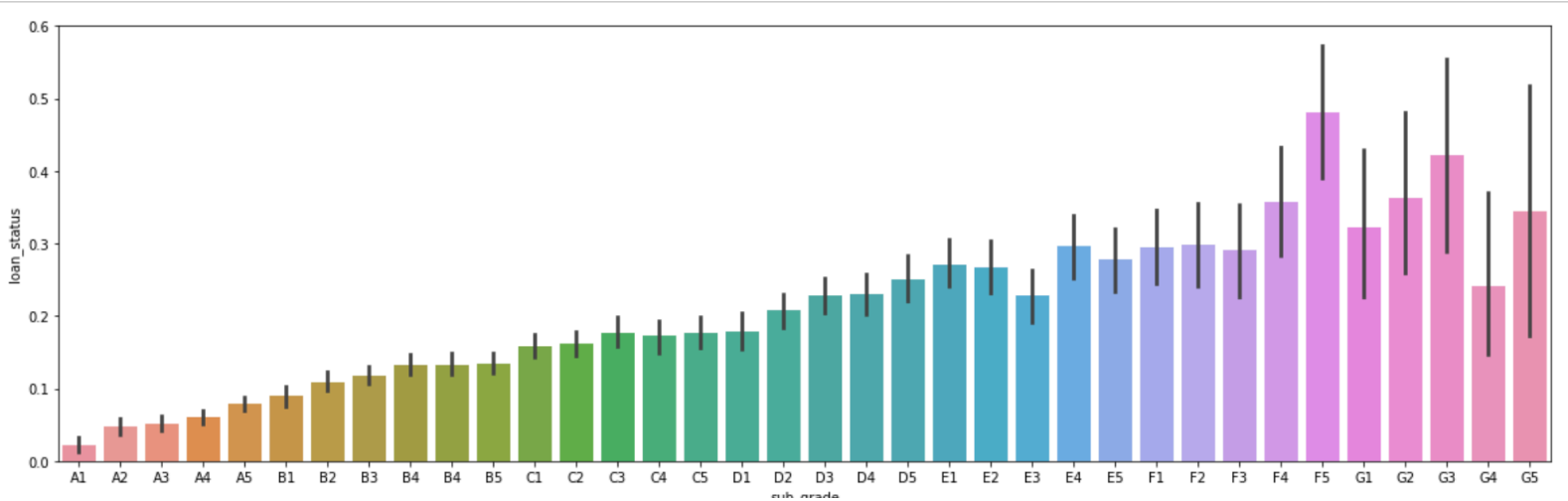
As the Interest range icreases loan defaults increases as well

EDA Analysis

Univariate analysis

5. Grade:





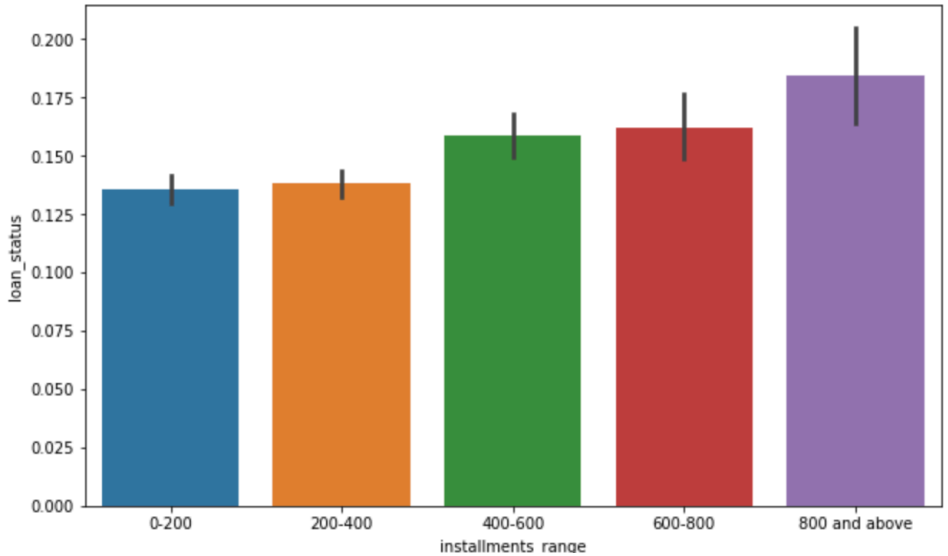
Observation

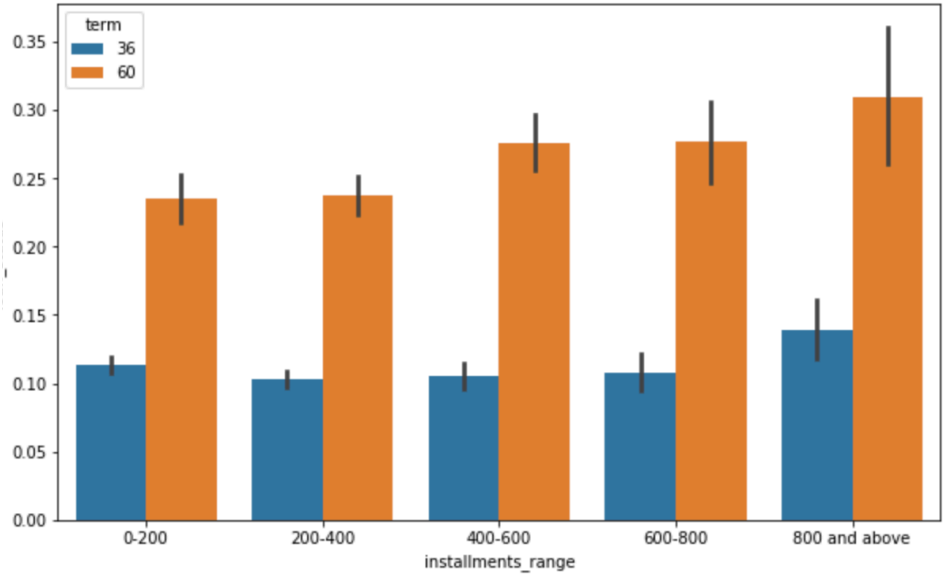
As the grade changes from A to G the default rate increased as well, this makes sense because "G" grade loans are riskier than that of "A"
The sub grade, they follow the same trend except at E3, G1 and G4

EDA Analysis

Univariate analysis

6. Instalments:





Observation

As the installemts increase the default rate increases as well.

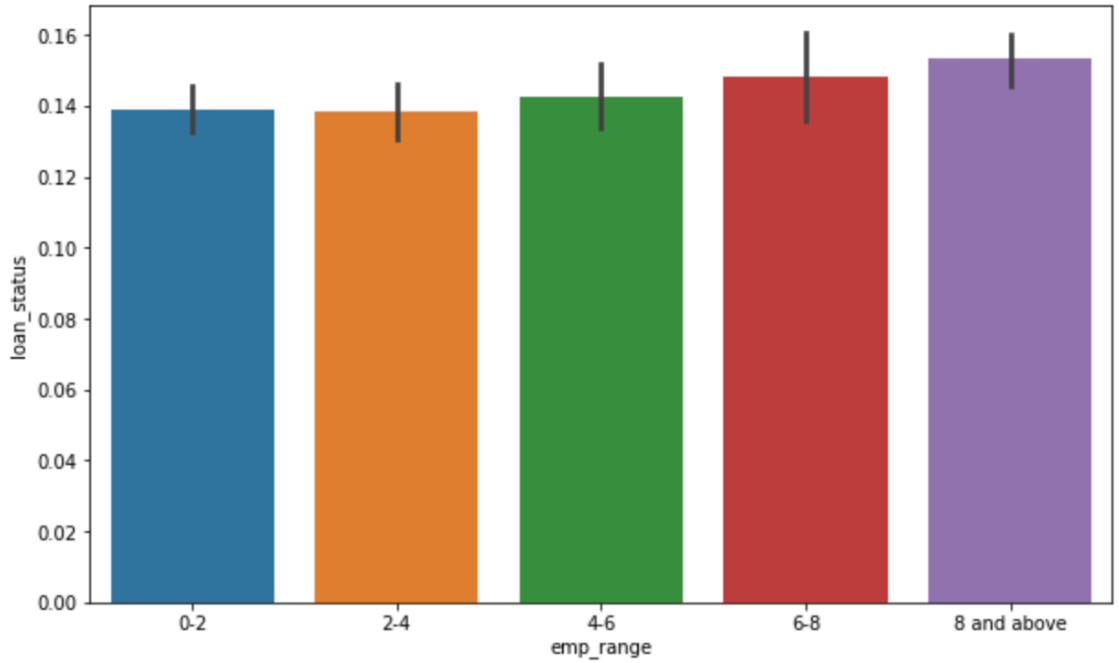
Also if you observe the default rate doesn't change much at 36 term level but at 60 term level it gradualllly increase

so longer term loans with high installemnts are likely to default more

EDA Analysis

Univariate analysis

7. Employment experience and Profession



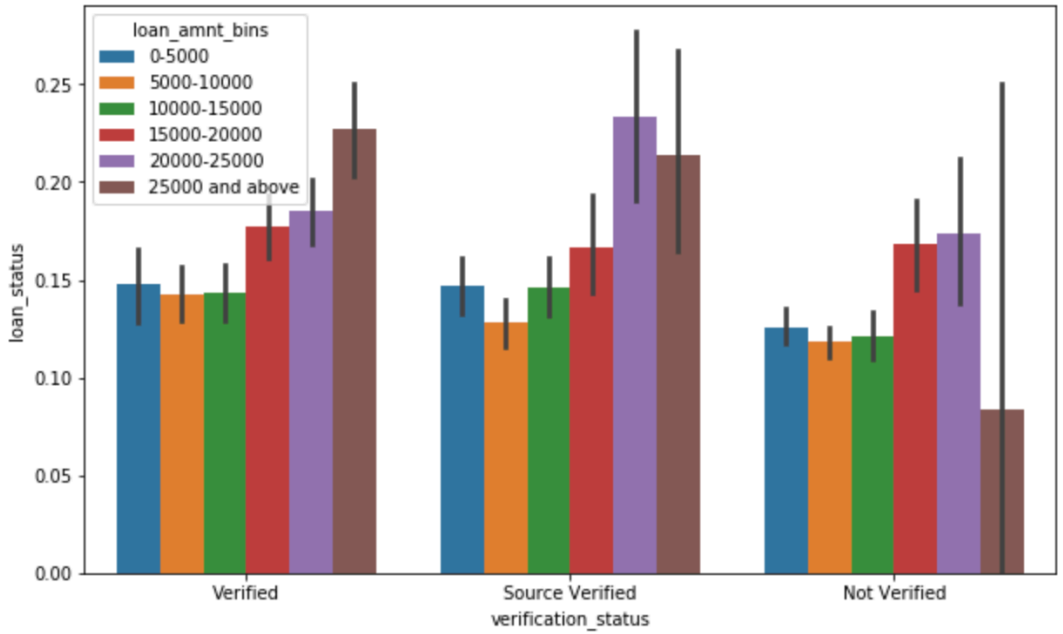
Observation

There isn't much variation in emp_length and there are 27,982 unique values in emp_title so it doesn't help either. So Default rate, isn't much affected by these variables

EDA Analysis

Univariate analysis

8.Verification status :



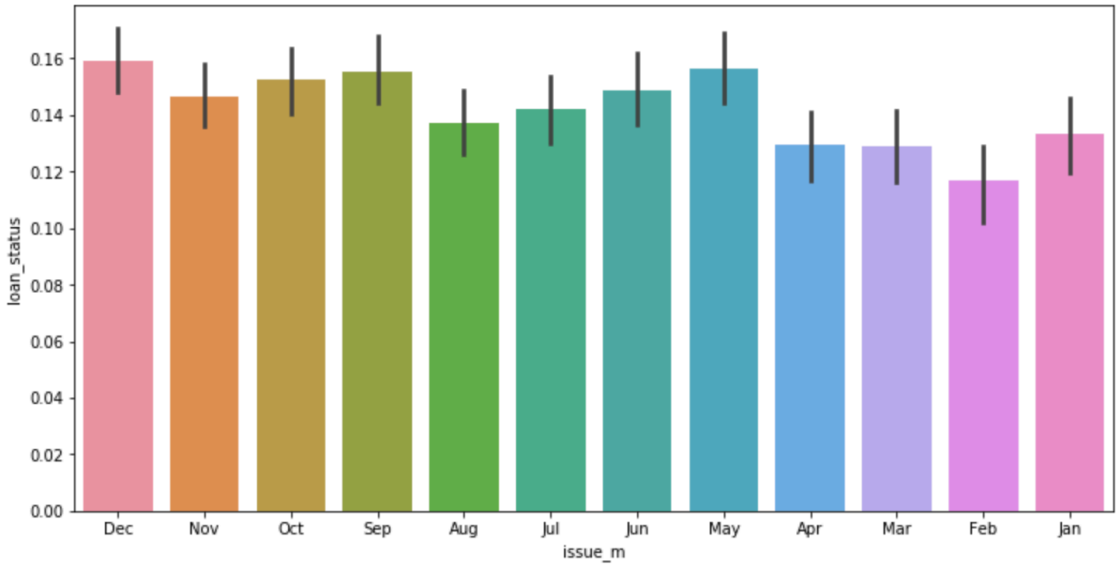
Observation

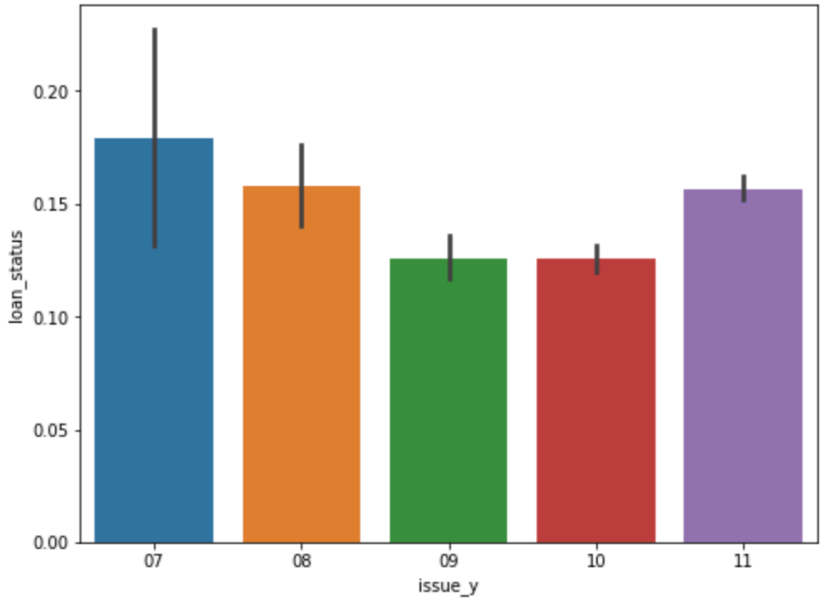
Verified loans have higher defalut rate, this shouldn't be the case in general, so have segmented it further based on loan amounts.
Higher loan amounts are Verified more often, and high loan amounts only have high default rates - so this justifies why verfied loans have high default percentage, that is because they contain maximum portion of high loans.

EDA Analysis

Univariate analysis

9. Issue month and year





Observation

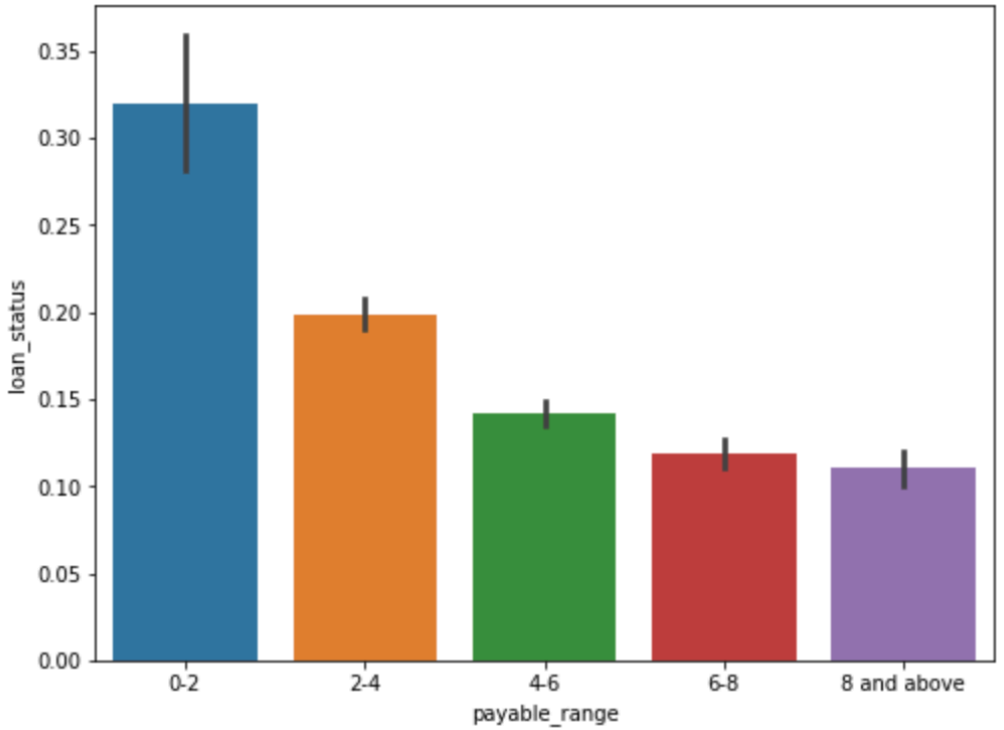
Not much help full, no variation month wise,
year wise the defaults have decreased upto 2009 same range upto 2010 and icreased again in 2011.

EDA Analysis

Segmented Univariate analysis

10. Payable ratio:

Annual income/loan amount



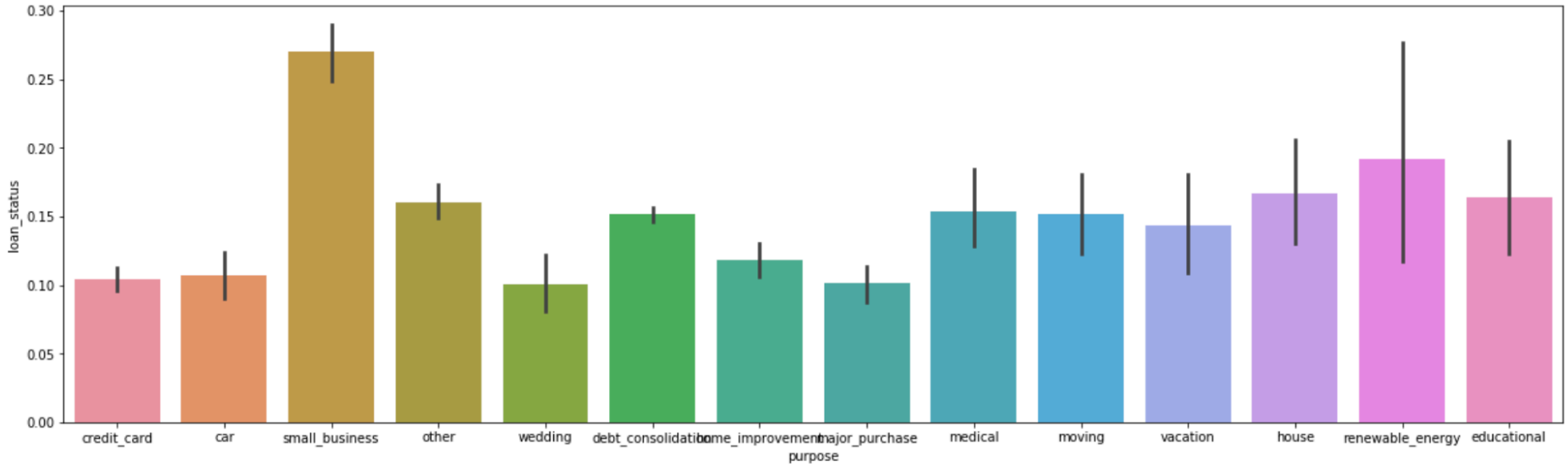
Observation

As expected people with low anual income and high loan amounts default more. This is a very good parameter

EDA Analysis

Univariate analysis

11. Purpose



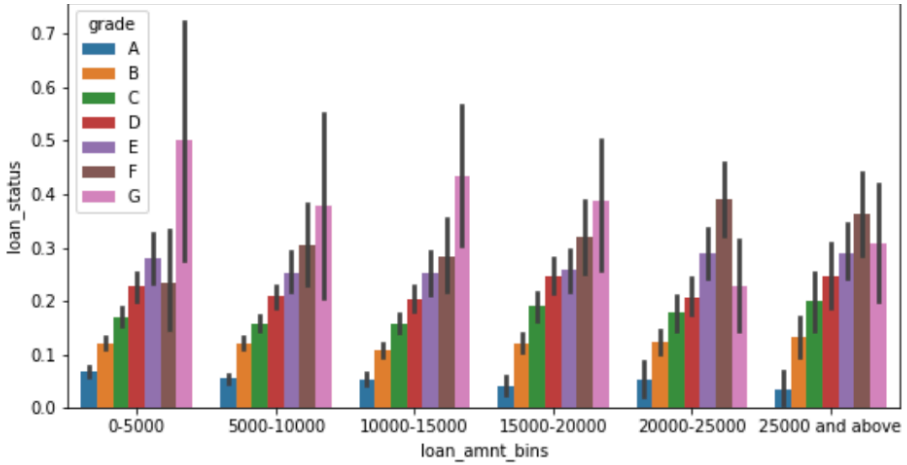
Observation

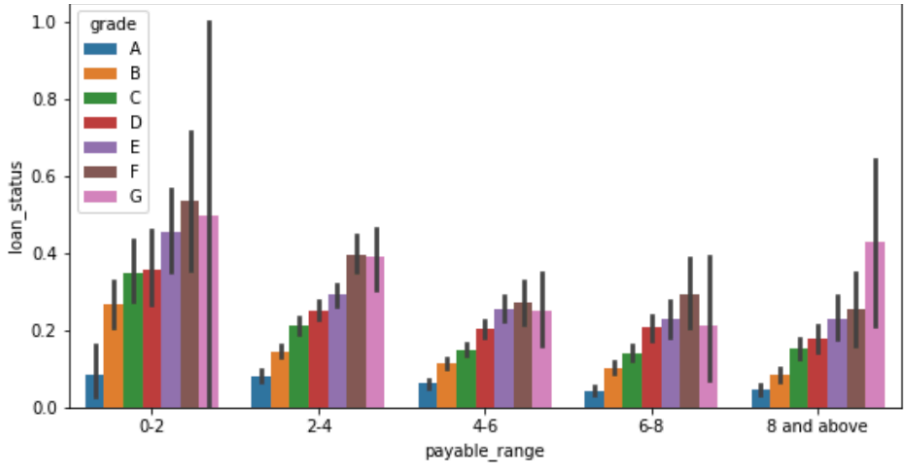
On an average loan for small businesses have high default rates, followed by renewable energy and house

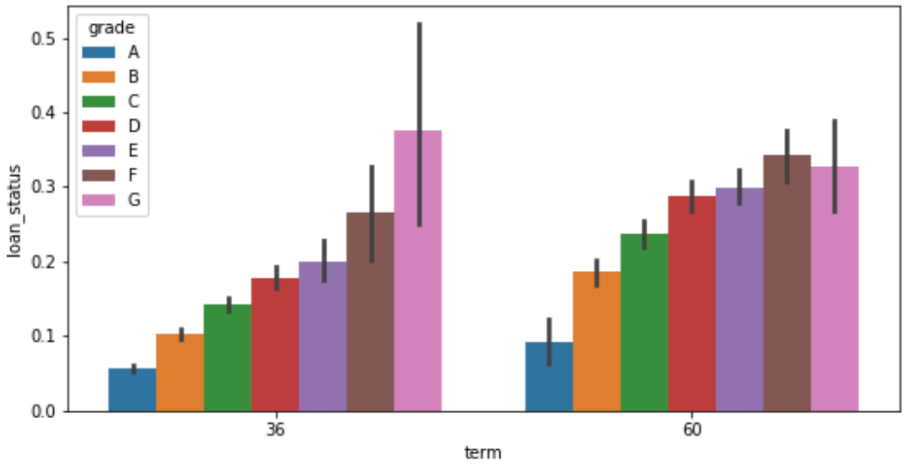
EDA Analysis

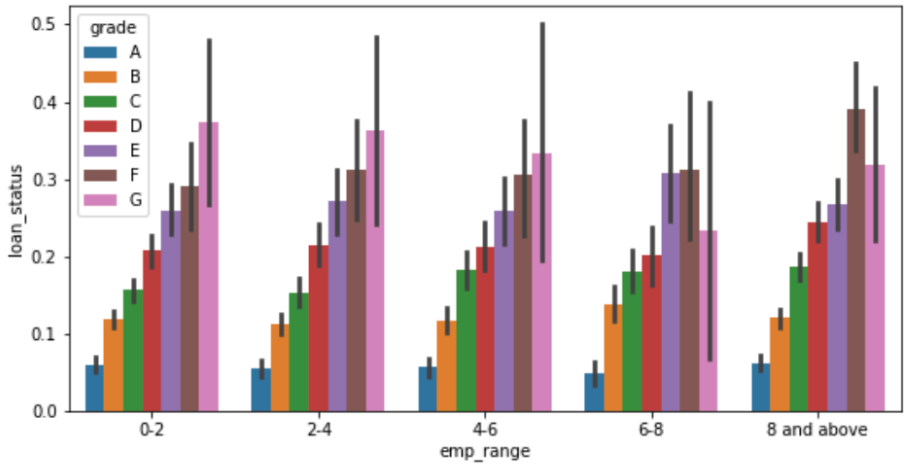
Segmented Univariate analysis

1. Grade





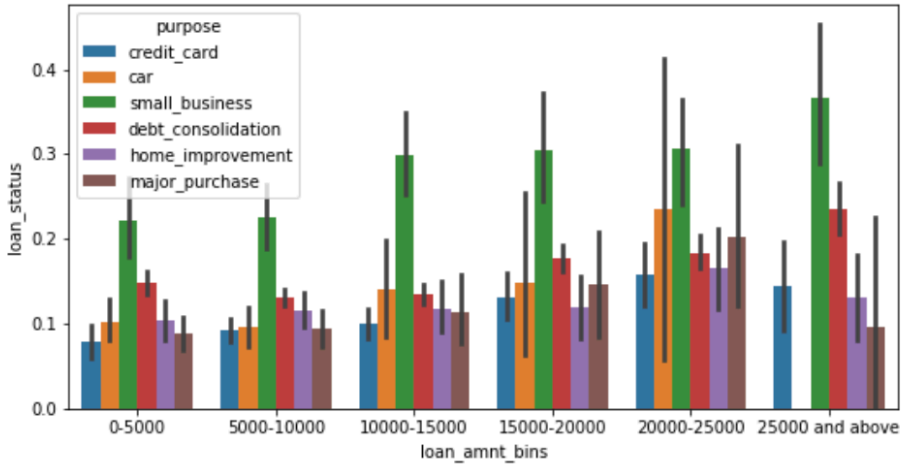


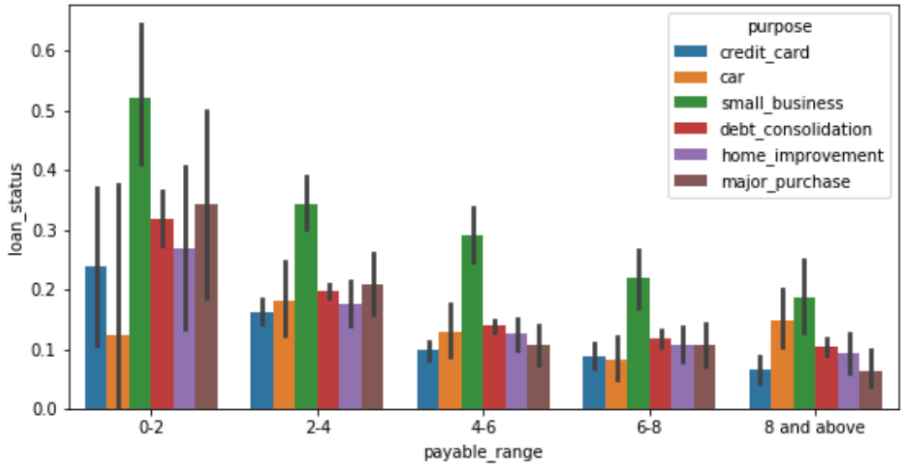


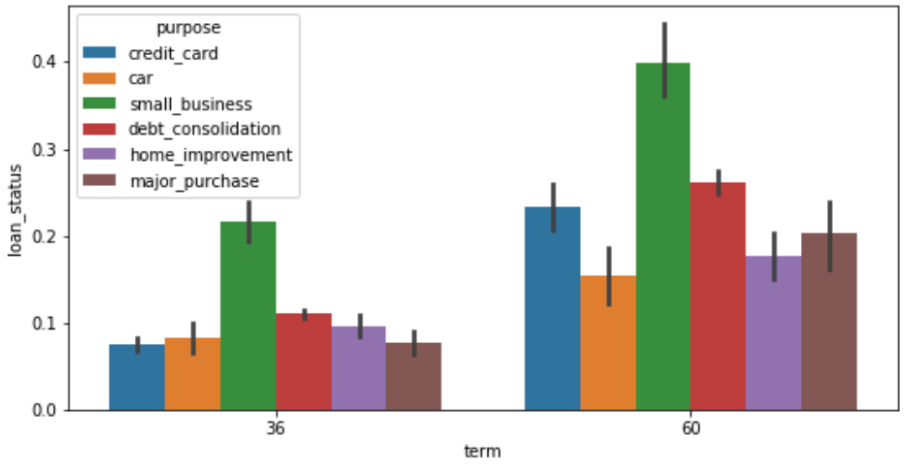
EDA Analysis

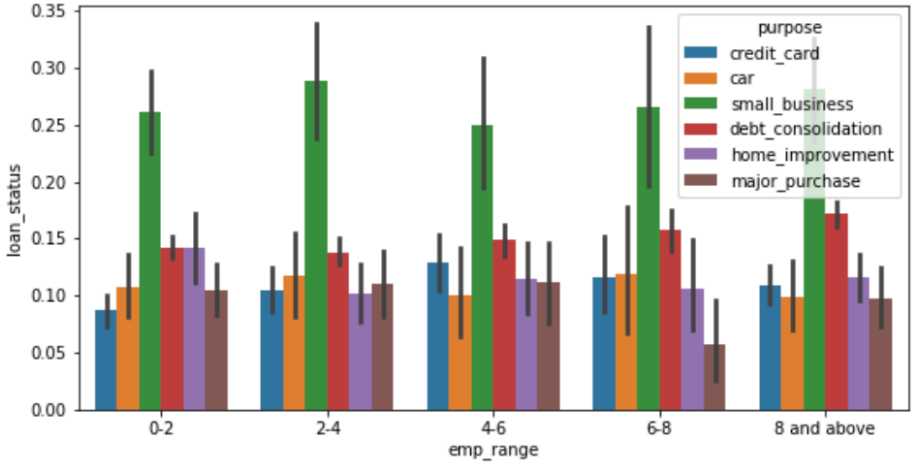
Segmented Univariate analysis

2. Purpose





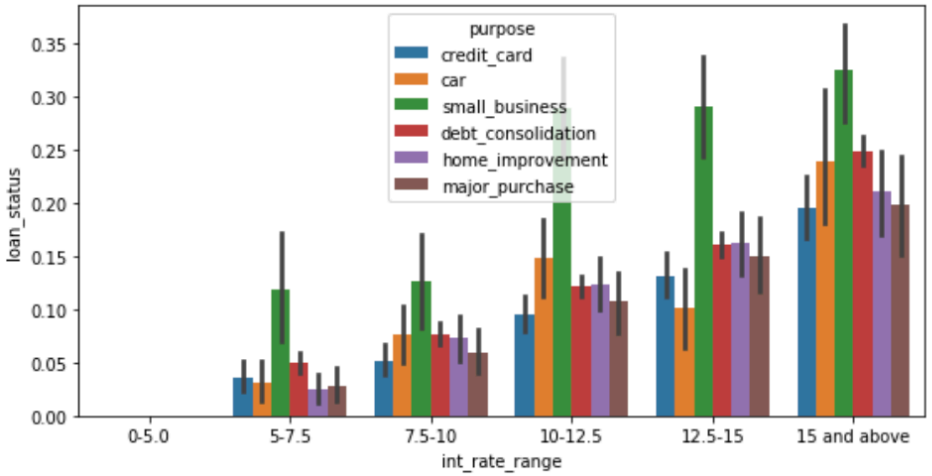


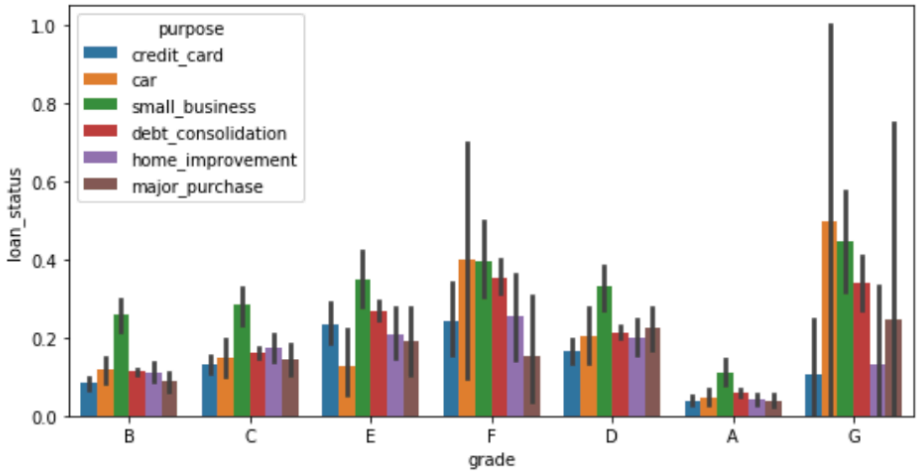


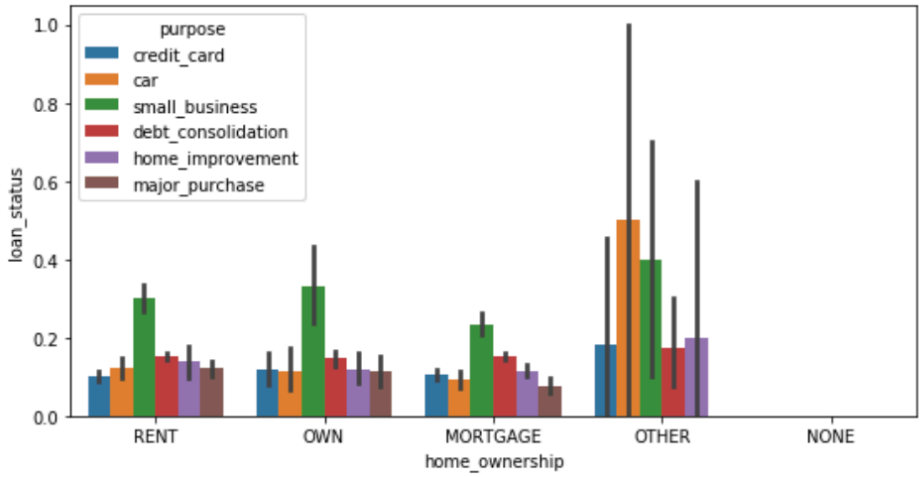
EDA Analysis

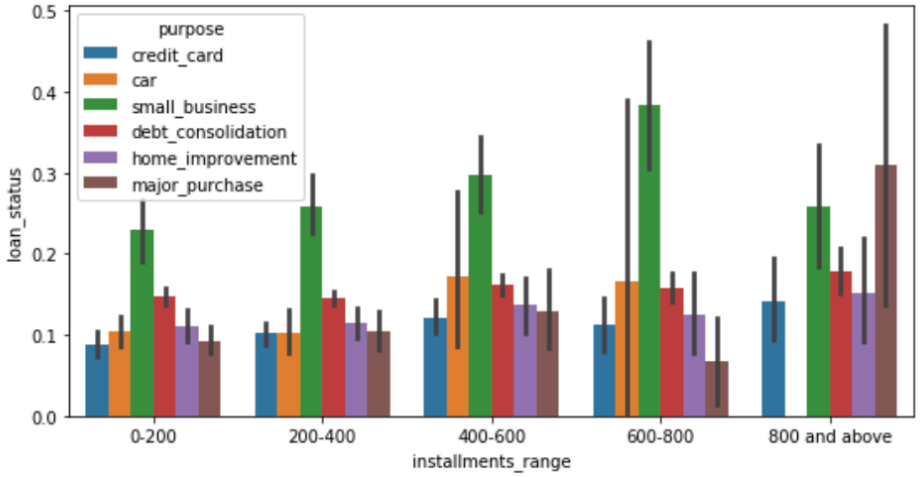
Segmented Univariate analysis

2. Purpose





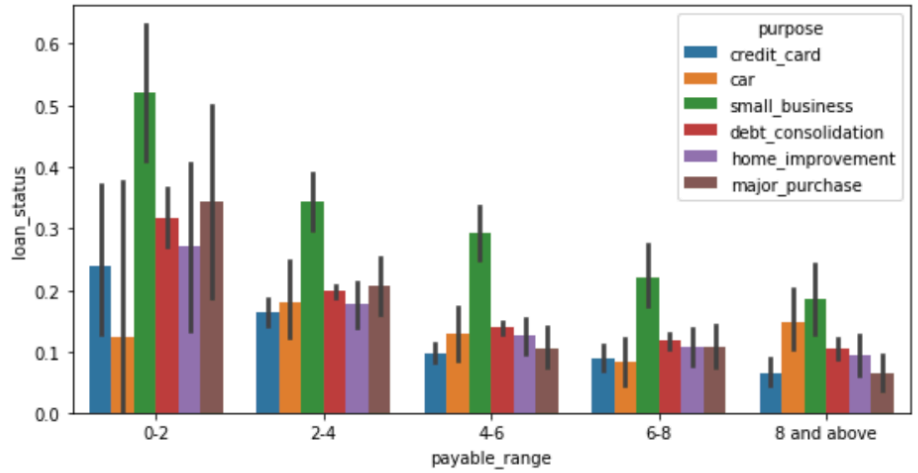


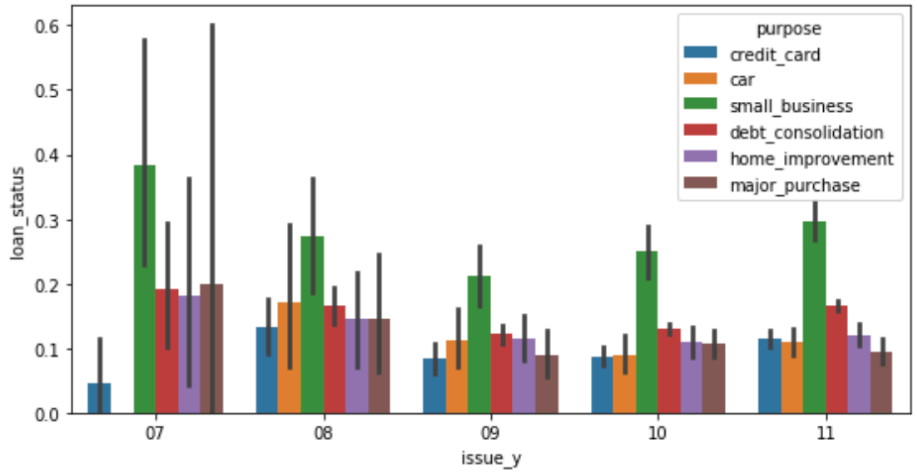


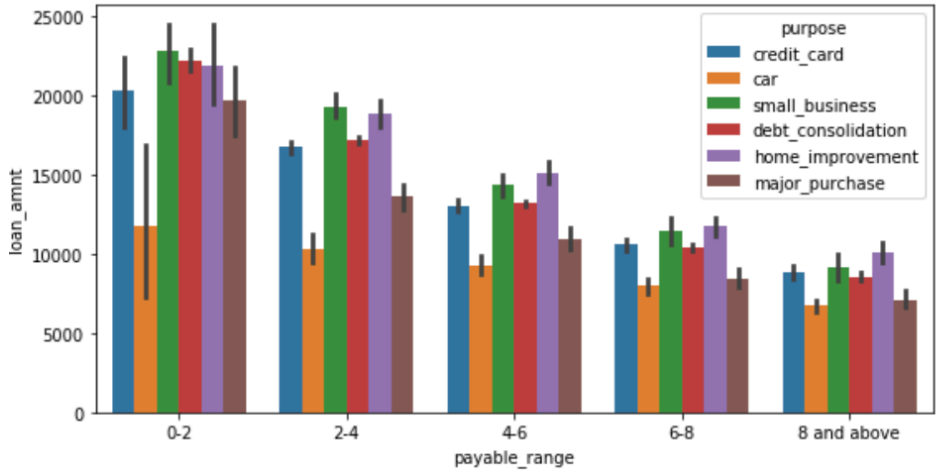
EDA Analysis

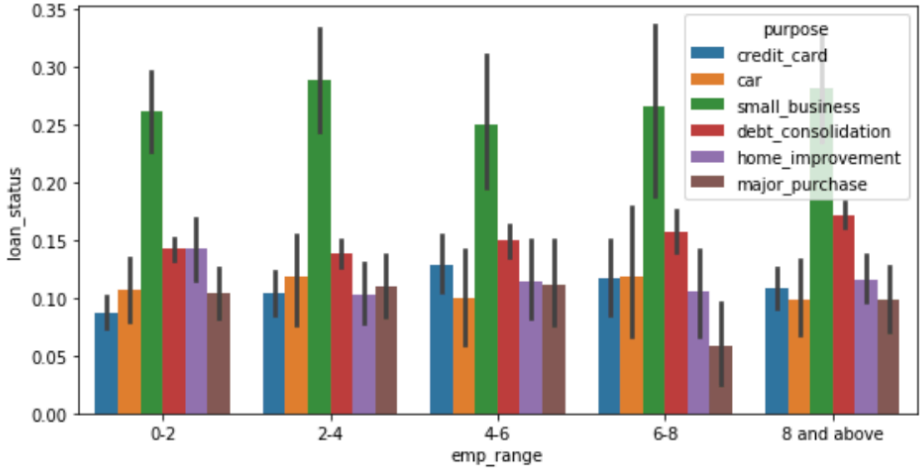
Segmented Univariate analysis

2. Purpose









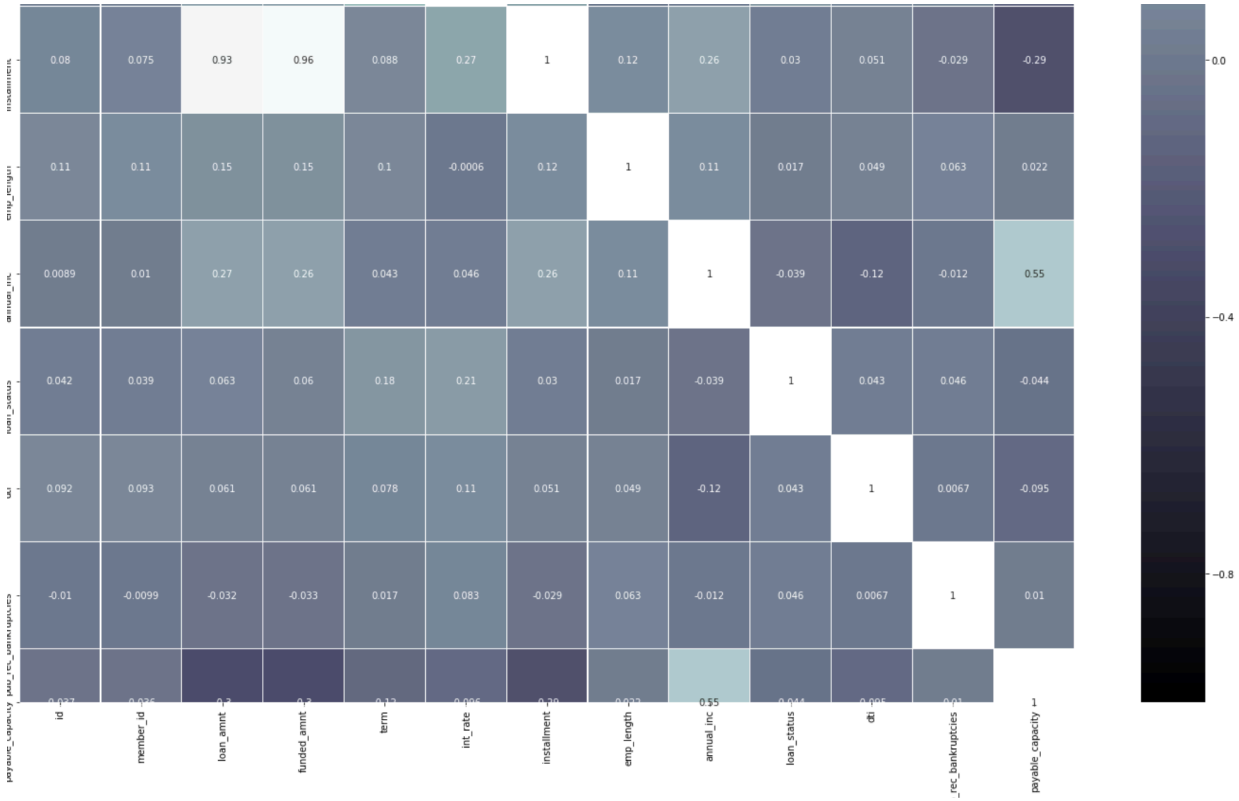
EDA Analysis

Bivariate analysis

Heat map

Observation

+vely correlated variables are
1.loan amount, funded amount and installment
2.loan amount and Interst rate, interest rate and term
Higher these variables higher is the default rate
Annual income and Payable capacity are +vely corelated
-vely correlated variables are
Anual income, payable capacity are very less correlated with most of the other variable.
-vely corelated with loan status, so higher these values lower is the default risk of default.



Conclusion

1. Loan_ amnt: You can notice apart from first bin, as the loan amount increases the default rate increases as well
2. Funded amount invested: follows the same pattern as the loan amount, that is, as the funded amount increases so does the default rate
   It simply meams that higher loan amounts are likely to get charged of compared to lower loan amounts
3. Term: Longer the loan term higher is the default rate
4. Interest rate: As the Interest range icreases loan defaults increases as well
5. Grade: As the grade changes from A to G the default rate increased as well, this makes sense because "G" grade loans are riskier than that of "A"
   The sub grade, they follow the same trend except at E3, G1 and G4
6. Installemnt: As the installemts increase the default rate increases as well.
   Also if you observe the default rate doesn't change much at 36 term level but at 60 term level it graduallly increase
   so longer term loans with high installemnts are likely to default more
7. Emp lenght and title: There isn't much variation in emp_length and there are 27,982 unique values in emp_title
   so it doesn't help either. So Default rate, isn't much affected by these variables
8. Verification status: Verified loans have higher defalut rate, this shouldn't be the case in general, so have segmented it further based on loan amounts.
   Higher loan amounts are Verified more often, and high loan amounts only have high default rates - so this justifies why verfied loans have high default percentage, that is because they contain maximum portion of high loans.
9. Issue month and year: Not much help full, no variation month wise,
   year wise the defaults have decreased upto 2009 same range upto 2010 and icreased again in 2011.
10. Payable ratio (Annual Income/loan ammount): As expected people with low anual income and high loan amounts default more. This is a very good parameter
11. Purpose: On an average loan for small businesses have high default rates, followed by renewable energy and house