Lending Club Case Study

SUBMISSION

Group Members:

- Naina Jain
- Thejasvini V

Lending Club Case Study

Problem statement and Business Objective

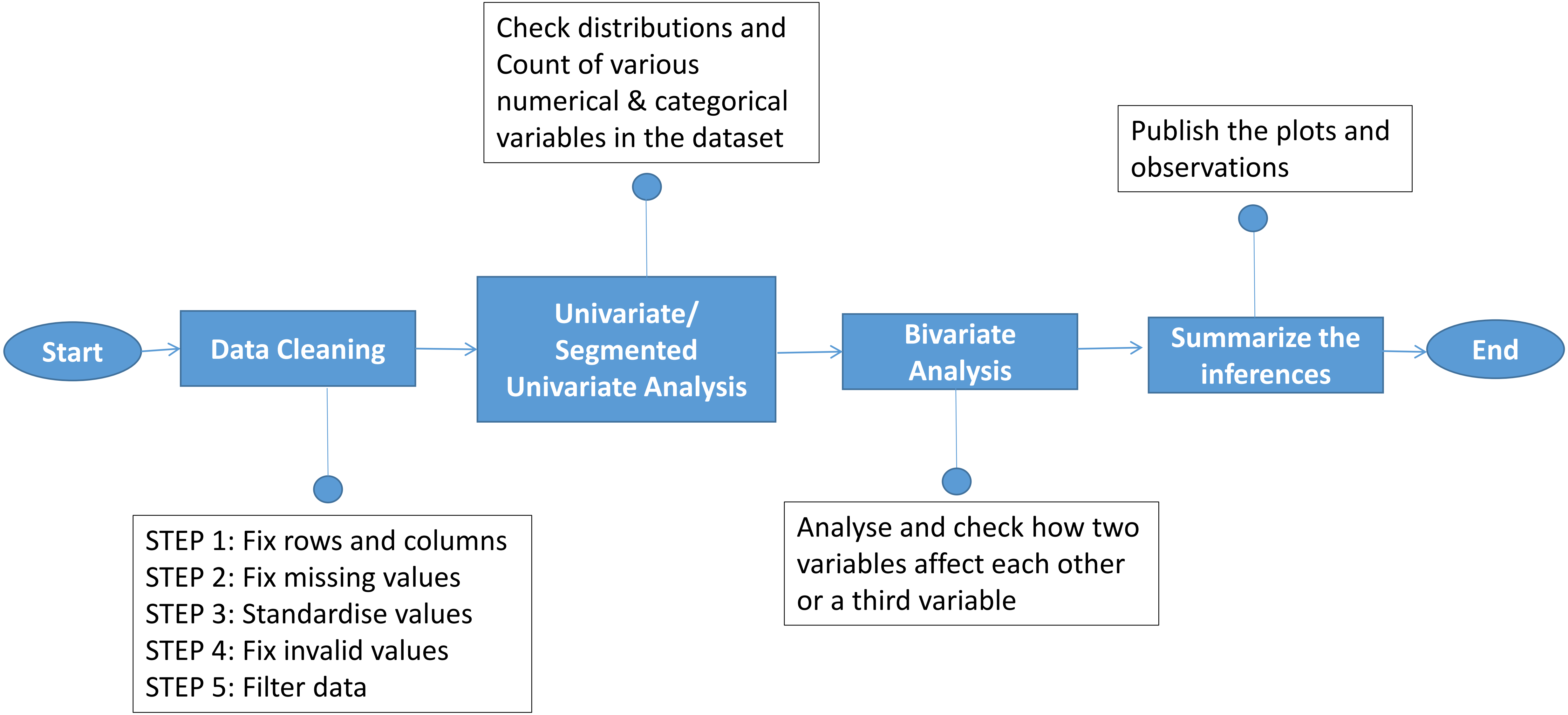
Lending Club (LC) is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures. Borrowers can easily access lower interest rate loans through a fast online interface.

Objective is to understand the driving factors behind loan default, i.e. the driver variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.

To achieve this,

- Analyse the Loan data set of all loans issued through the time period 2007 to 2011.
- Data consists of Current, Fully Paid and Defaulted (Charged-Off) Loans.
- Analyse the data set using EDA to understand how consumer attributes and loan attributes influence the tendency of default

Problem solving methodology- Flowchart

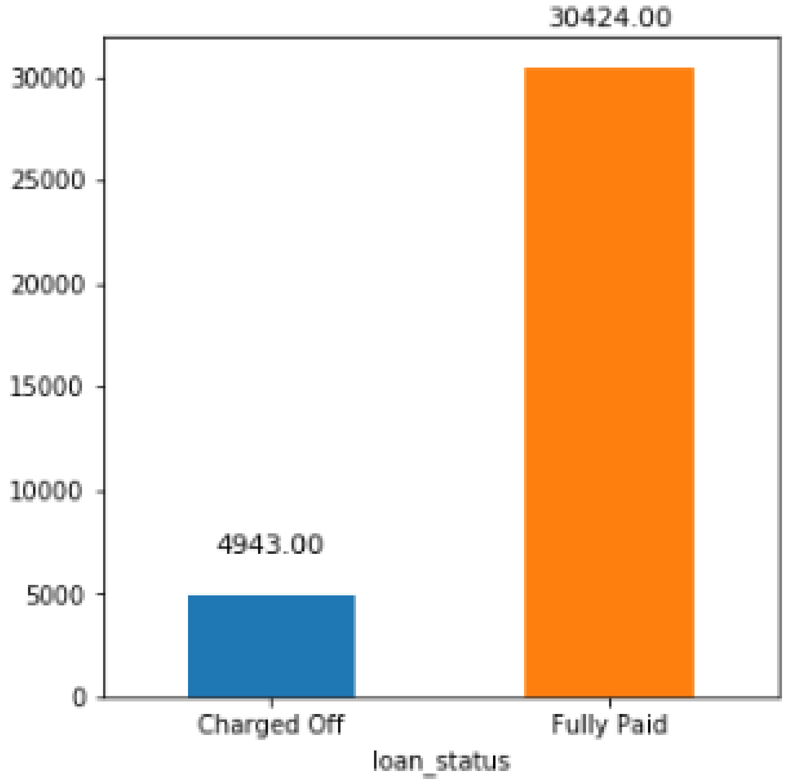


Data Understanding & Cleaning

- Understanding the column meanings based on data dictionary.
- Identifying the unnecessary columns which have no use in analyzing the loan dataset.
- Removing all null values columns and rows.
- Converting columns to relevant datatypes like removing % and converting date columns to date datatype
- Deriving year, month from date columns
- Binning numeric values for easy analysis

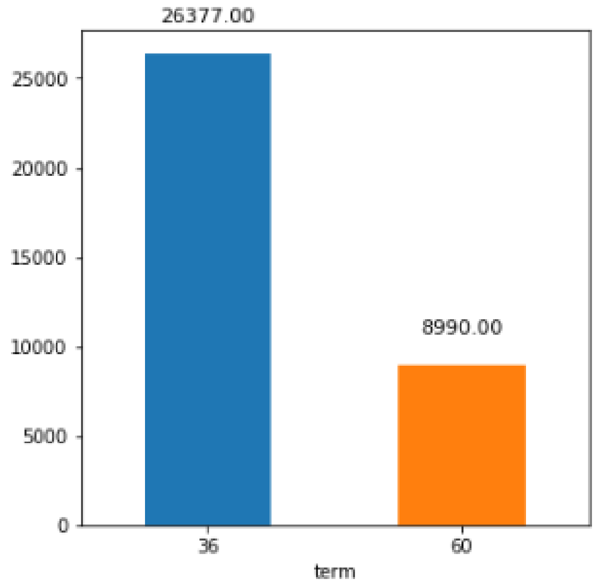
Understanding Loans based on “Loan Status”, “Term”

Approximately, 13.97 =~ 14% of the loans are Charged-Off (Defaulted)

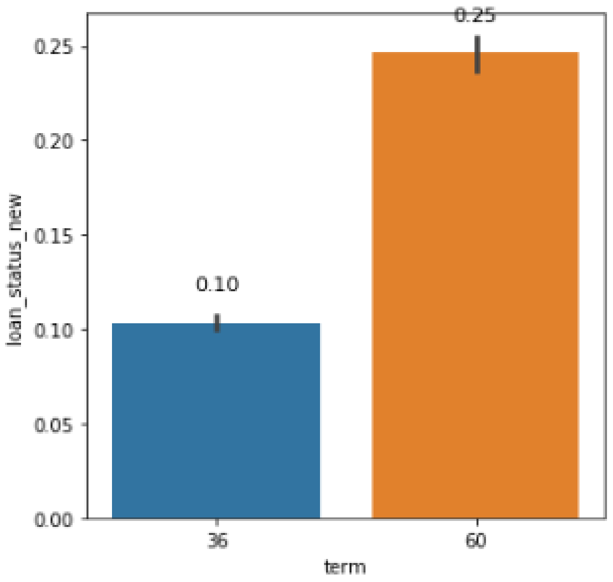


Loan status compared to term

Most of the loans are granted for '36 months'

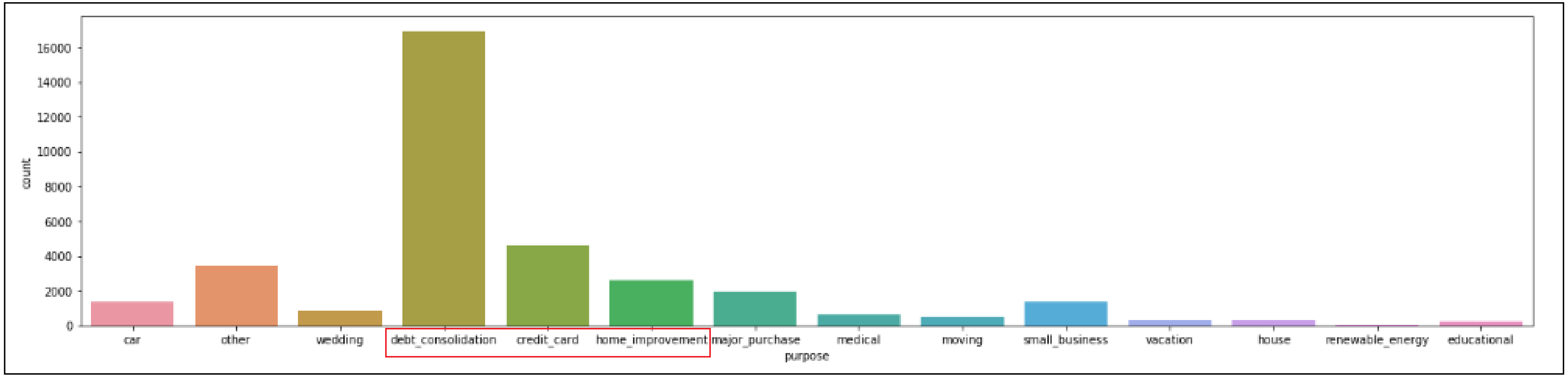


Loans granted for ‘60 months’ are more likely to default compared to ‘36 month’ loans

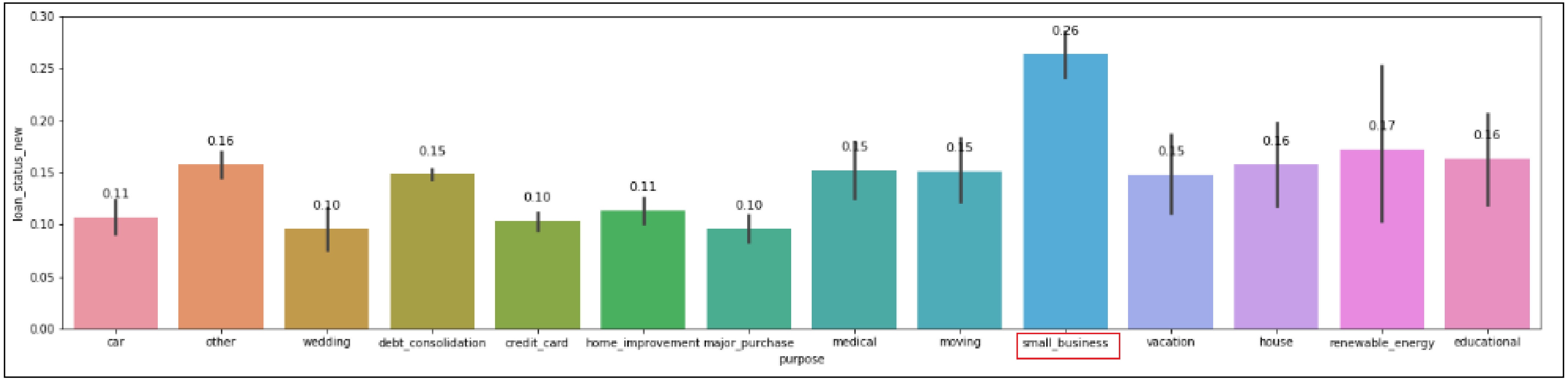


Understanding Loans based on “Purpose”

Majority of the loans are granted for ‘**debt_consolidation**’ followed by ‘**credit_card**’ and ‘**home_improvement**’



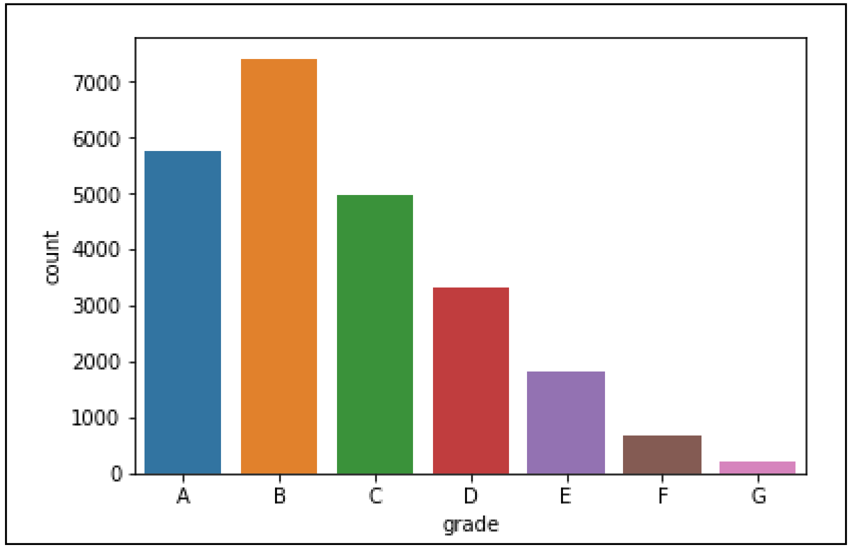
Loans granted for ‘**small_business**’ are likely to default more

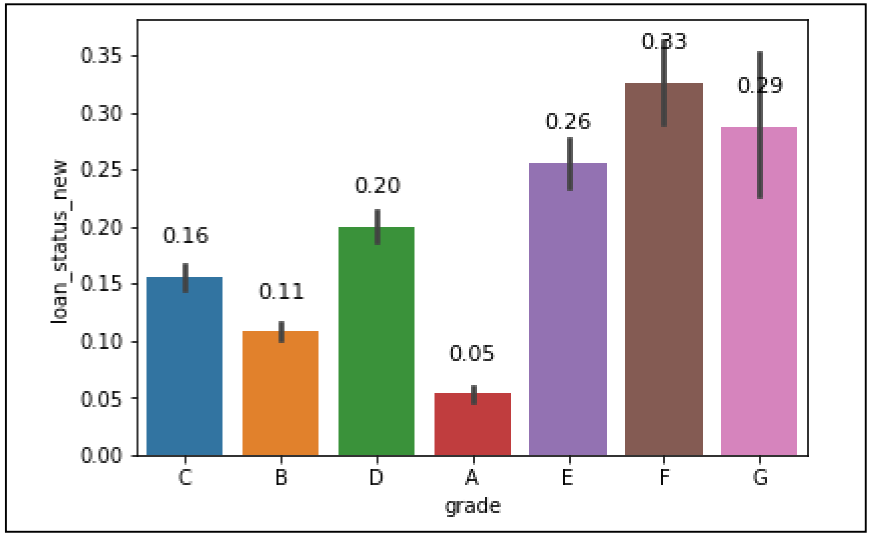


Understanding Loans and default based on “Grade”

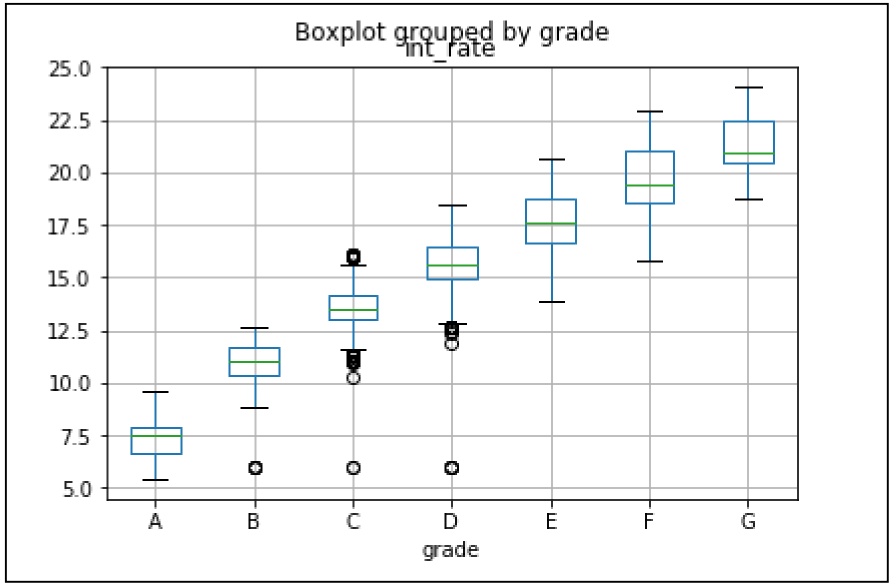
Lower grades (D,E,F,G) default more compared to high quality grades (A,B,C)

Most of the loans granted are of high quality with grades- ‘A’, ‘B’ and ‘C’

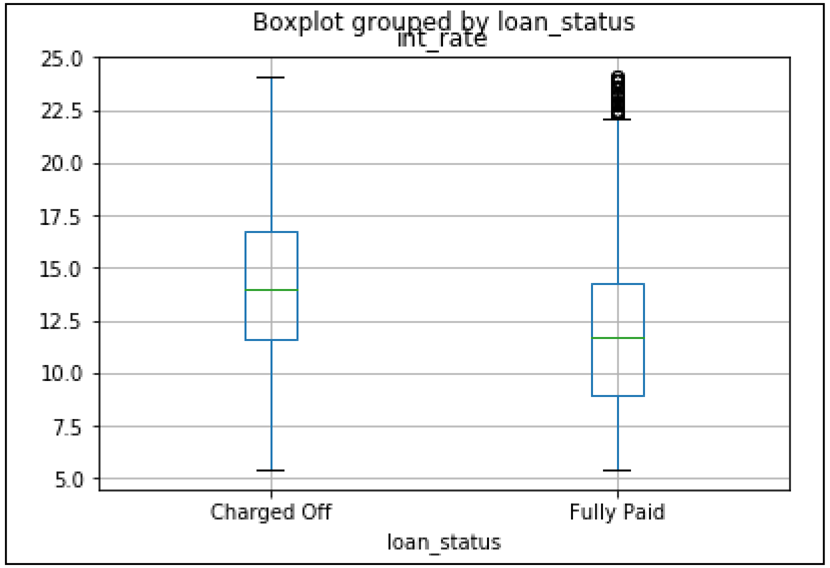




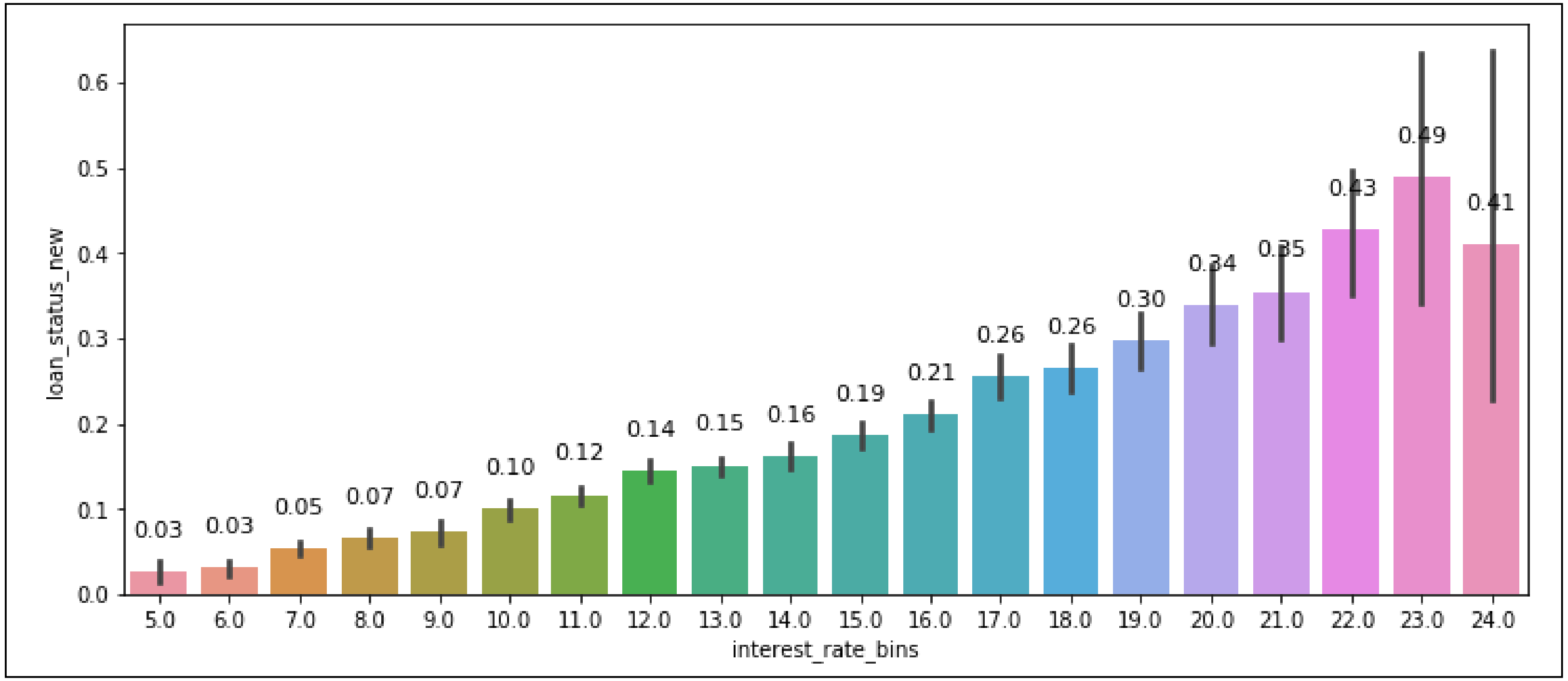
Interest rate varies directly with the grade. Larger or worst the grade, higher are the rates of interest.



Understanding Defaults based on “Interest rate”



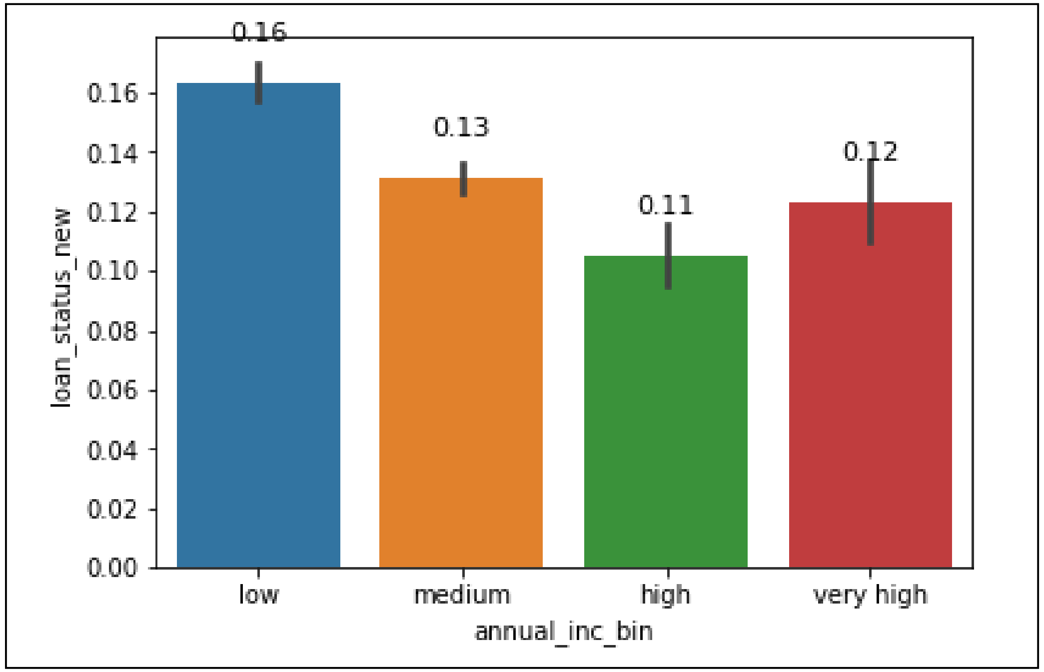
As the interest rate increases, the likelihood of default also increases.



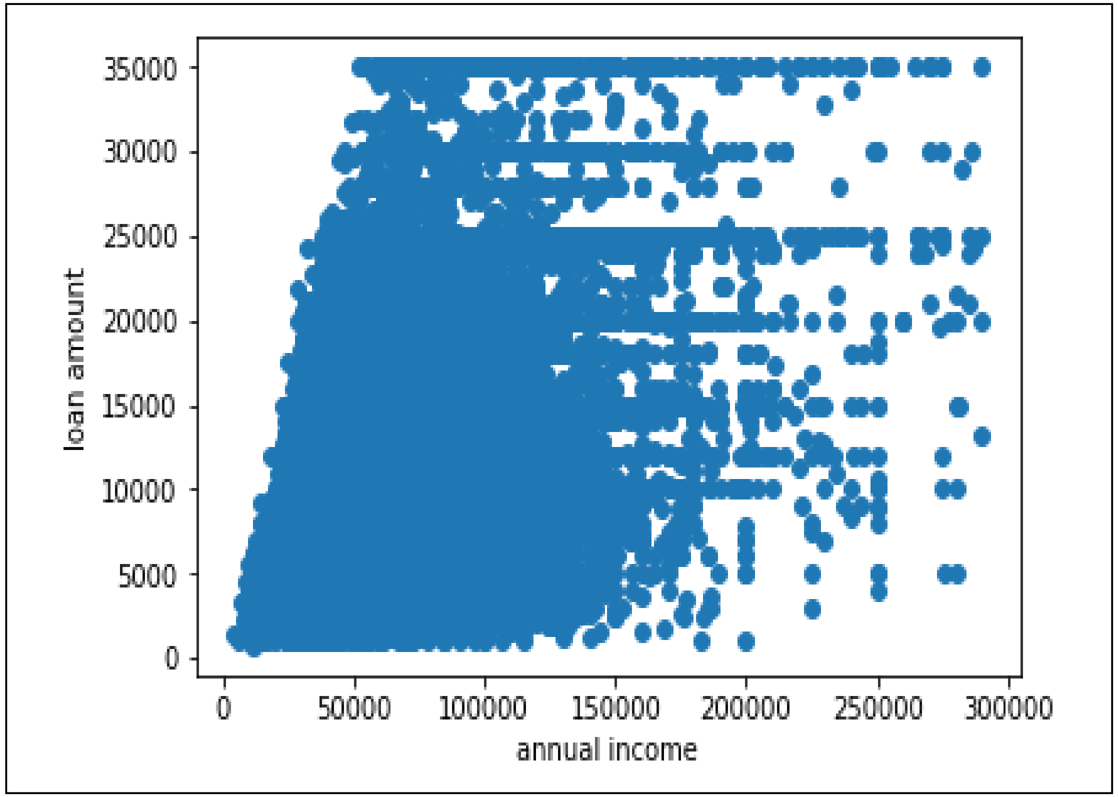
Understanding Defaults based on "Annual income"

Borrowers with lower annual income are more likely to default more. In below graph,

Low- annual income < 50000
Medium- annual income between 50000 and 100000
High- annual income between 100000 and 150000
Very High- annual income > 150000

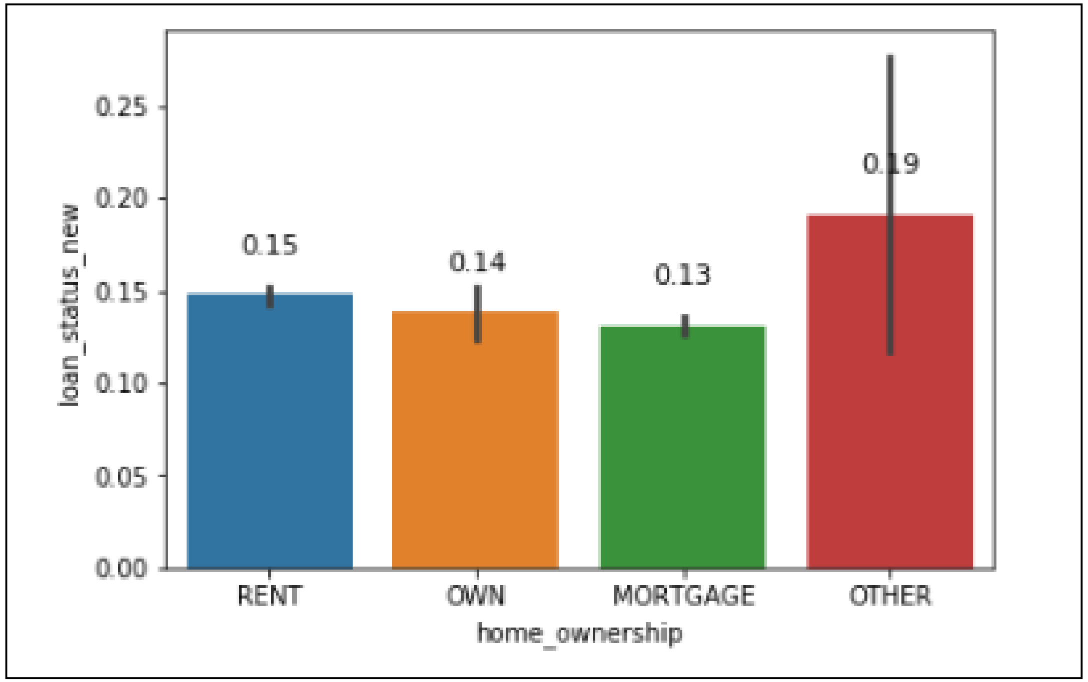


Borrowers with income less than 50000 are taking loan of high amount (>25000). This seems to be risky and likelihood of default



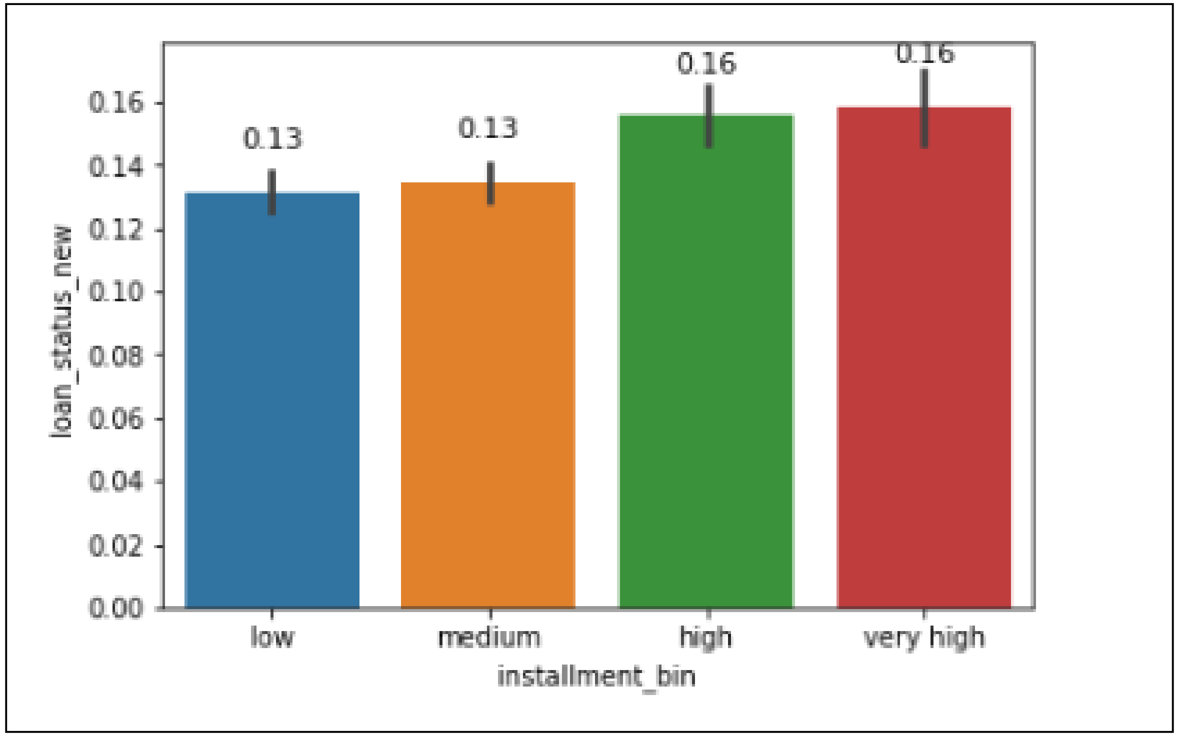
Understanding Defaults based on “Home Ownership”, “Installment”

'Other' is likely to default more compared to other values such as Rent, Own and Mortgage.



Higher the installment amount, the higher the default rate.
In below graph,

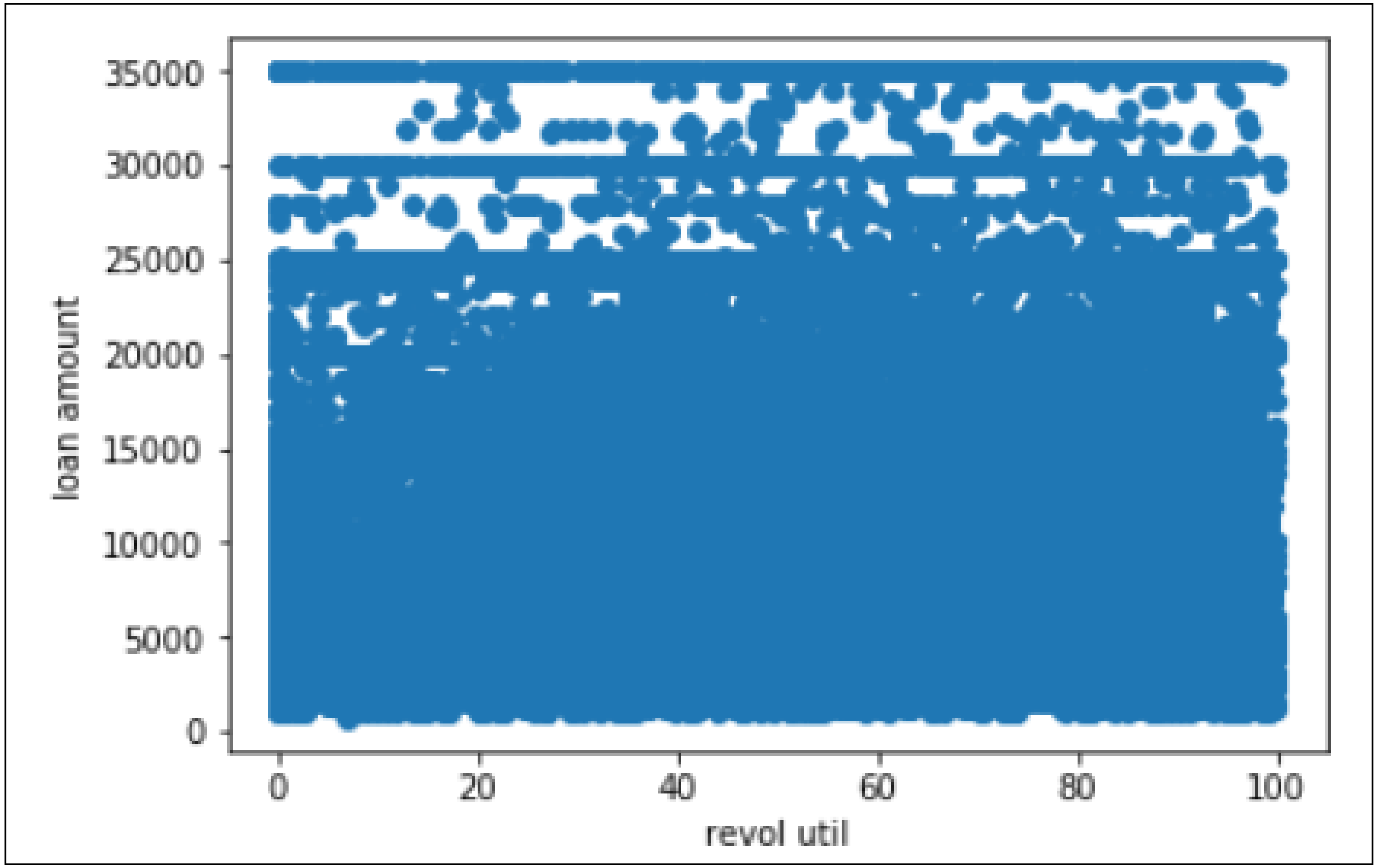
Low- installment <= 200
Medium- installment> 200 and <= 400
High- installment >400 and <=600
Very High- installment >600



Understanding Defaults based on "Revolving line utilization rate"

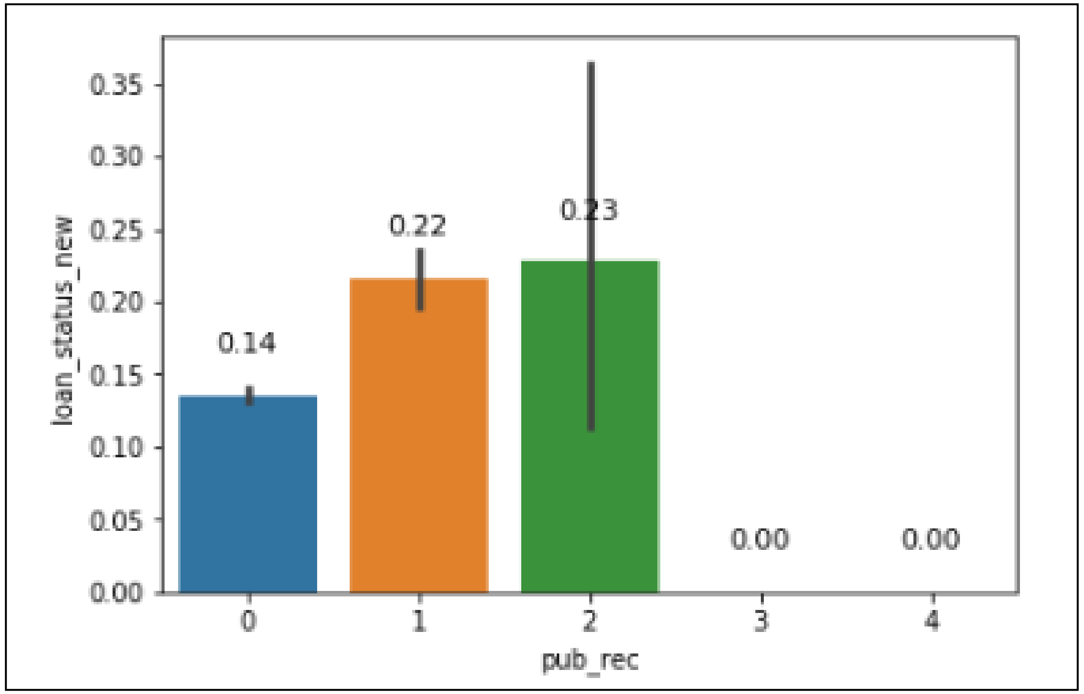
There have been some high value loans extended to borrowers with revolving line utilization rate of higher than 75%.

Density of low value loans is also high. They should be approved less often.

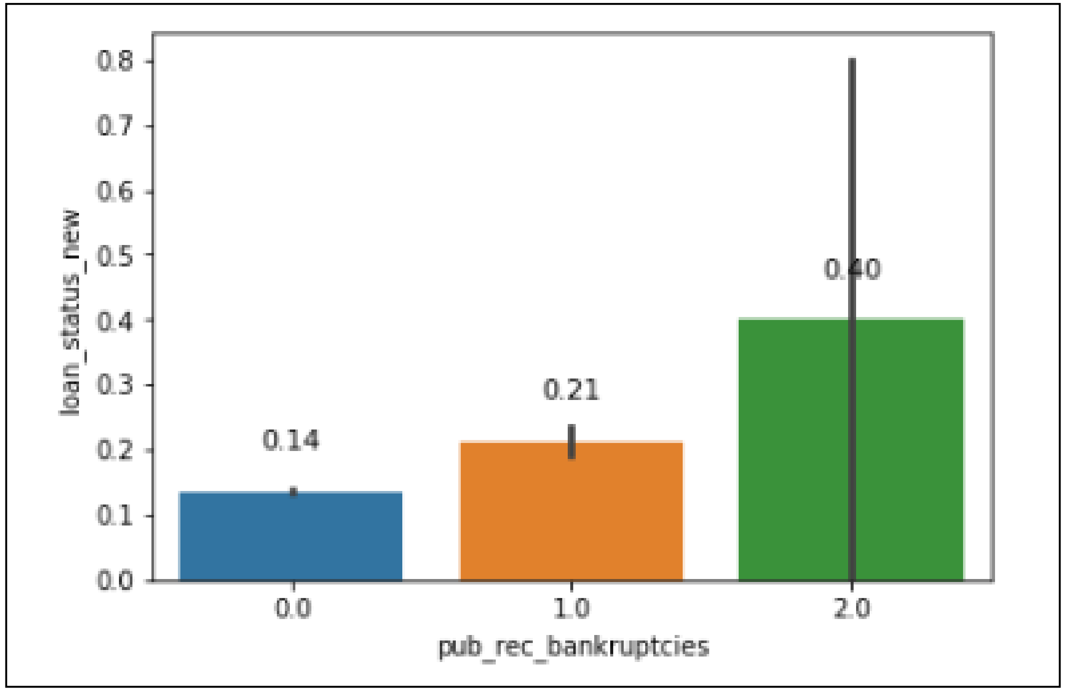


Understanding Defaults based on “pub_rec” and “pub_rec_bankruptcies”

Having even 1 or more public derogatory records increases the chances of defaulting

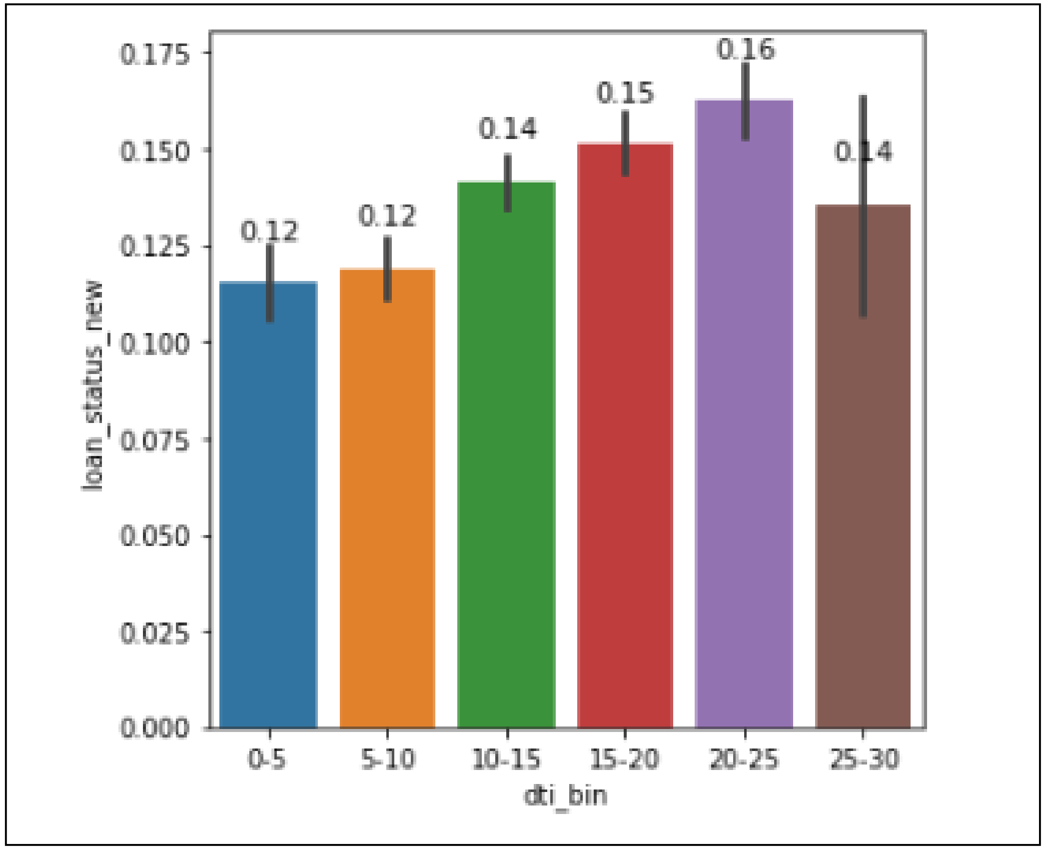


Higher the Number of public record bankruptcies, higher the likelihood of default

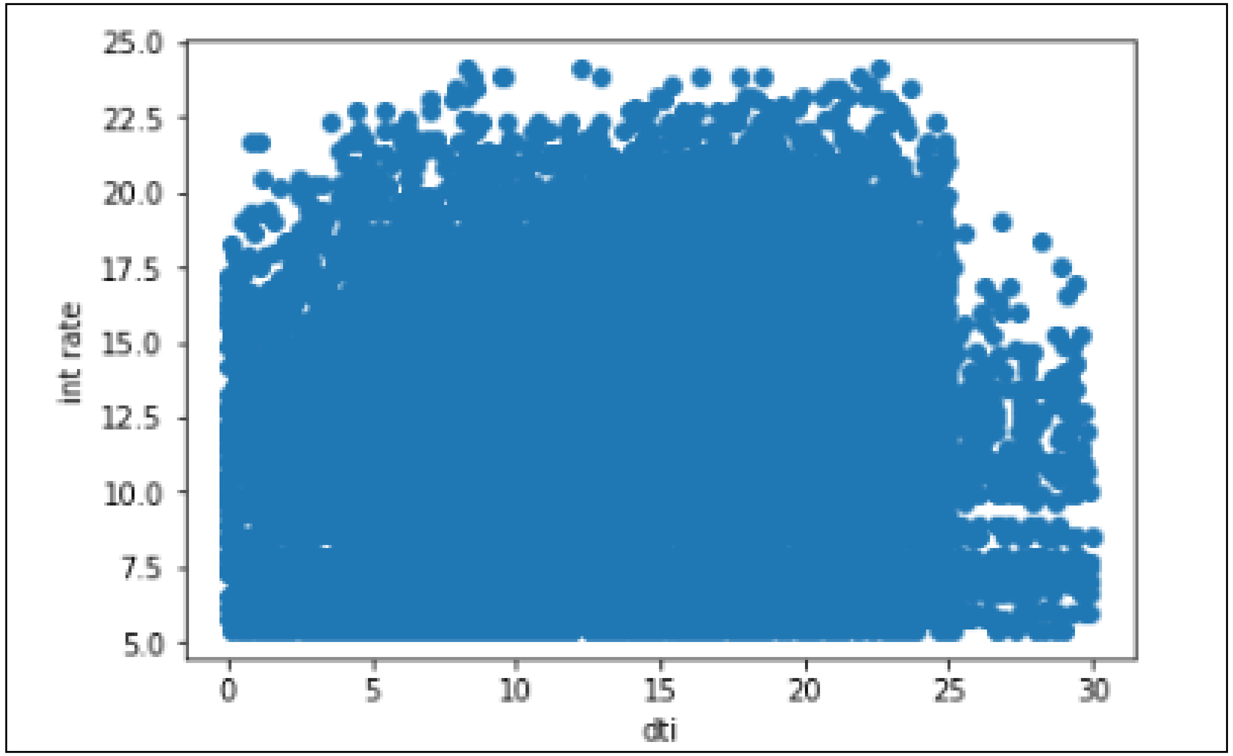


Understanding Defaults based on “Debt to Income Ratio”

Higher 'dti' signifies higher default rates. i.e. dti >20 have higher chances of default



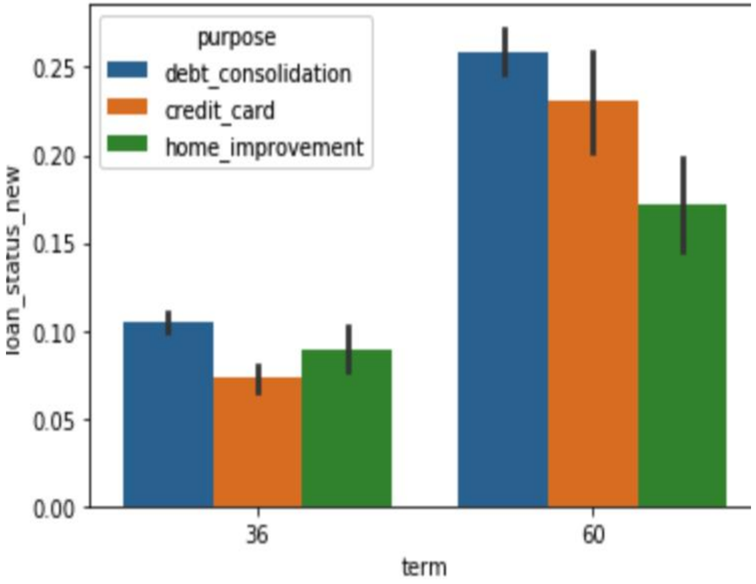
Higher interest rates should be charged for higher dti, but we see spread across all values without being differentiated in terms of interest rates

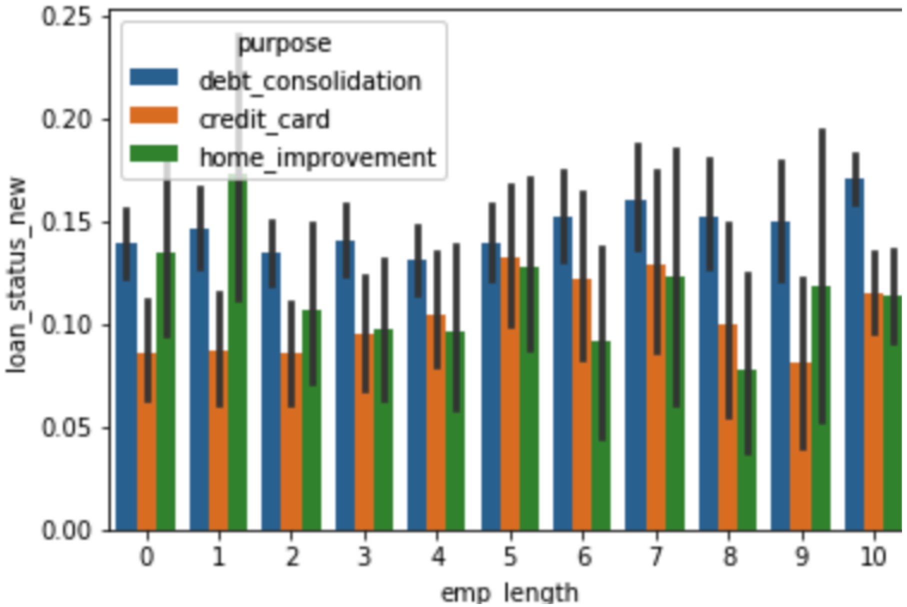


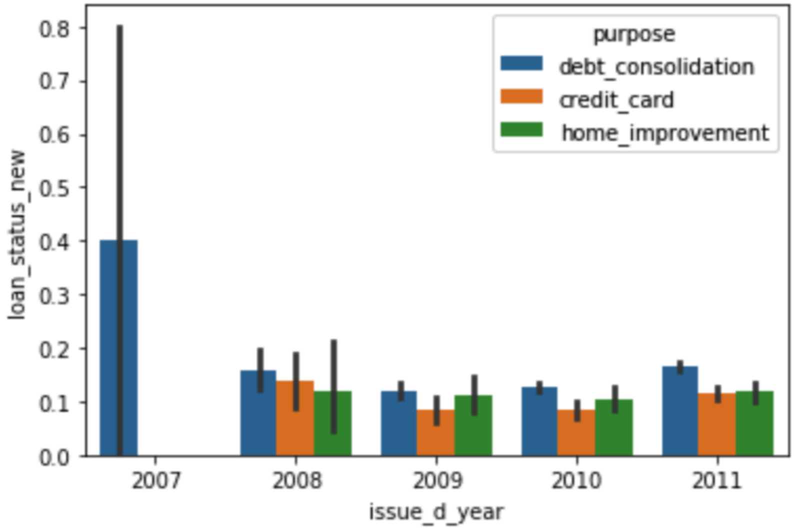
Analysis–Term/ InterestRate /emp length/ issuedate for top 3 purpose

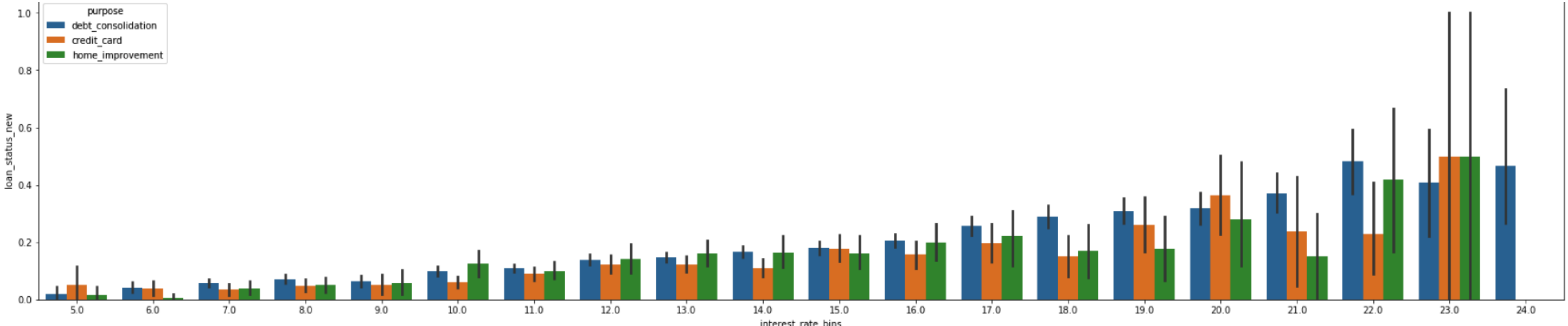
UpGrad

Clearly, debt_consolidation are more likely to default. Loans by employee for higher years of experience and taken for 60 months term having high interest rate is more likely to default.

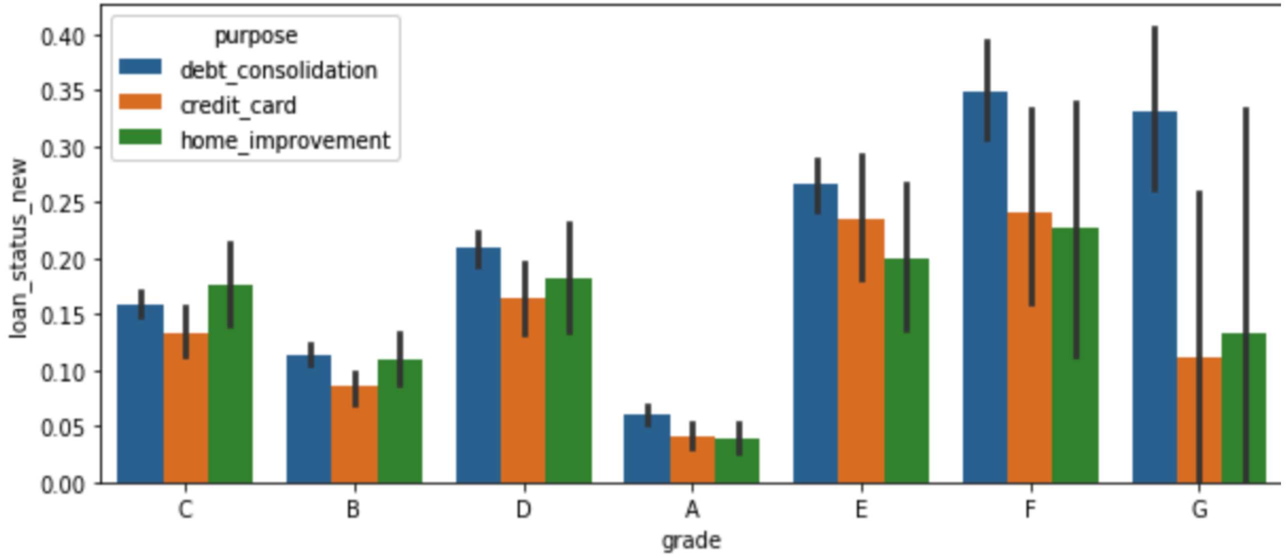


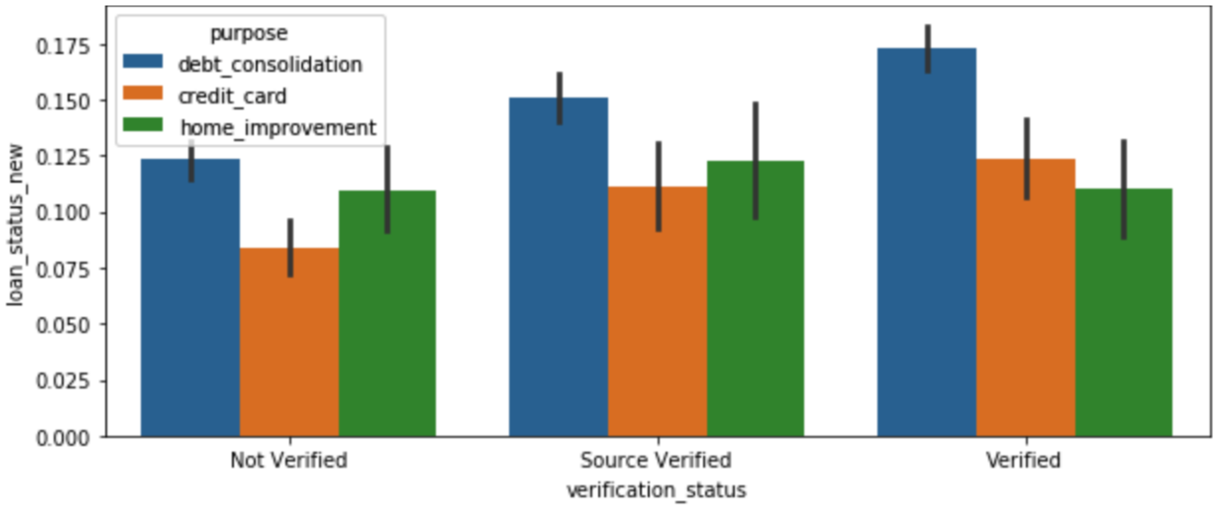






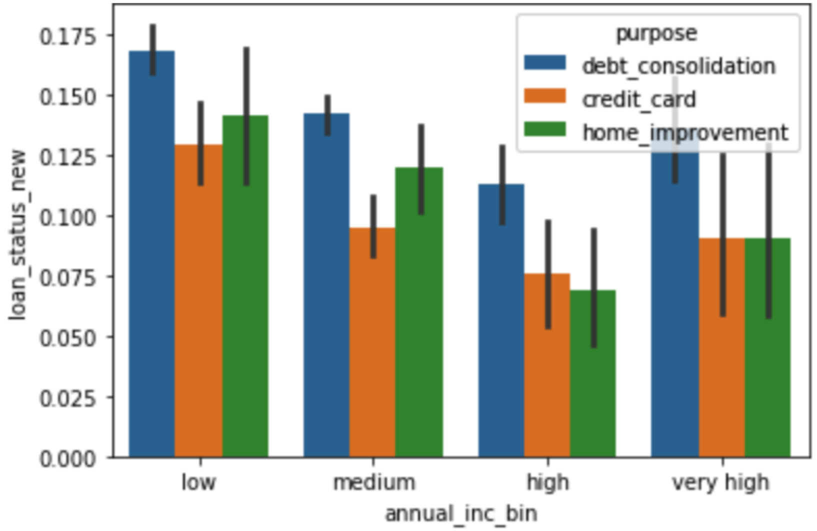
Analysis – Grade / Verification / annual Income / Installment for top 3 purpose

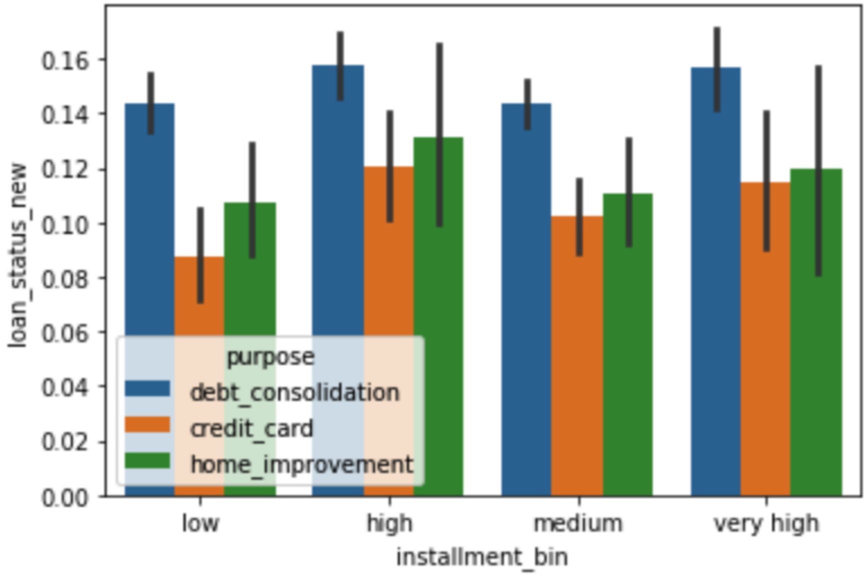




Clearly, debt_consolidation are more likely to default.

Loans having verified status with high instalments, low income and low grade are more likely to default.





Recommendations:

- ✓ Reduce the number of loans where term is ‘60 months’ as they are likely to default more compared to ‘36 months’.
- ✓ Reduce the number of approvals where the purpose of loan is “small_business”.
- ✓ Stop approving high amount loan to people with less annual income as they are likely to default more.
- ✓ Reduce approving loans to ‘Other’ under Home_ownership as they are likely to default more.
- ✓ Reduce number of approvals for loans where installment is greater than 400
- ✓ Stop approving loans where revolving line utilization rate is greater than 75%.
- ✓ Stop approving loans to people with previous bad record. Having even 1 bad record has high default probability.
- ✓ Charge higher interest rates for loans with dti greater than 15.