EDA CASE STUDY

SUBMISSION

BGNS Analytics:
1. Biswajit Dash
2. Sanyam Jain
3. Ganapathi Ramachandra Hegde
4. Beltur Rangarajan Navneet

Business Objective and Goal

❑Business Objective :

➢ To Identify the risky Loan applicants , who can potentially default the loan , to reduce Credit Loss to the company.

❑Data Analysis Goals :

➢ Identify the driving factors (variables which are string indicators of default).

➢ Identify the pattern of applicants likely to default the loan

➢ Recommend actions to reduce Credit Loss.

# Problem solving methodology – Flow Chart
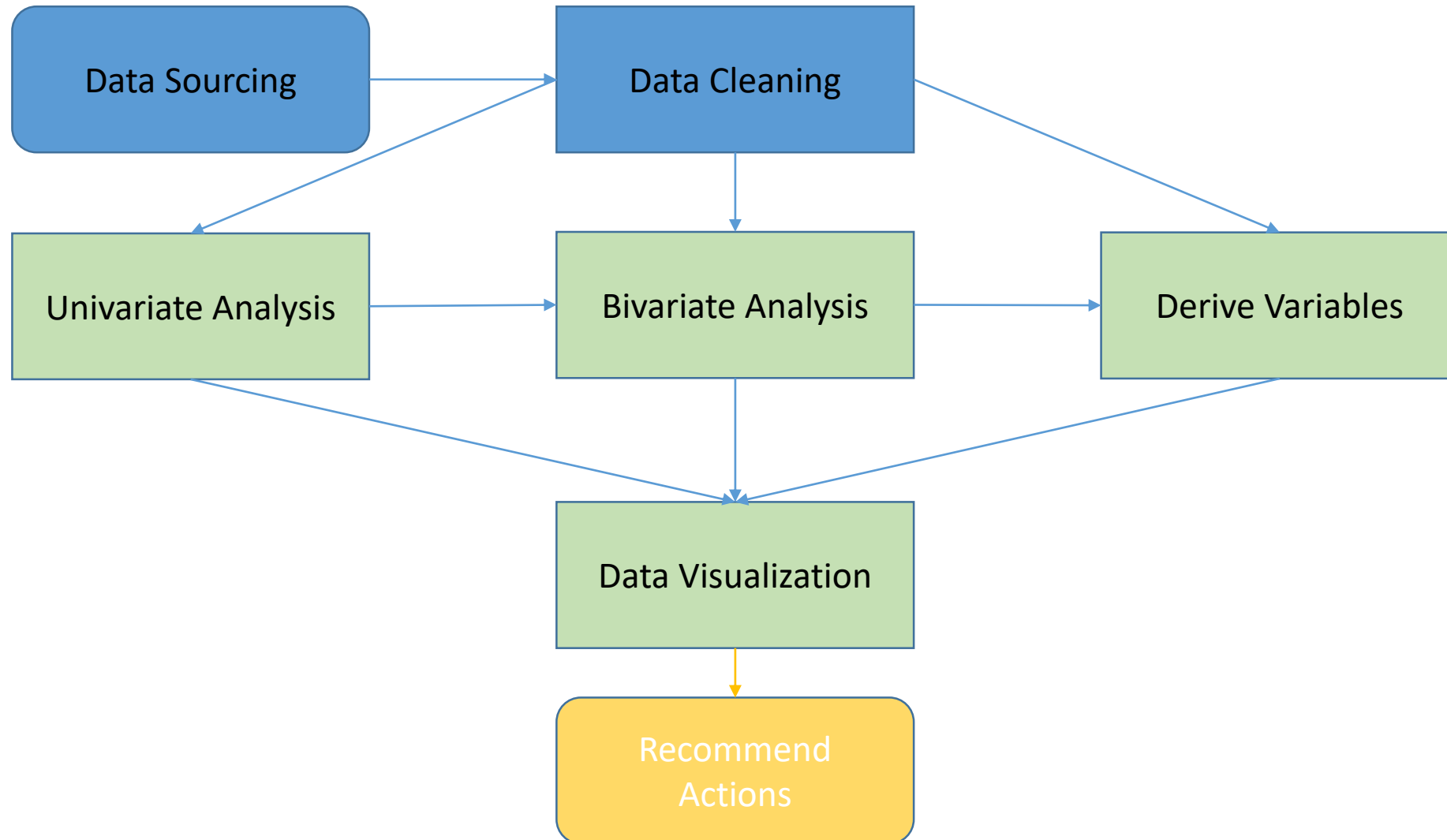
Data Cleaning

- Removed columns with all values set as NA/NULL.

- Removing columns with negligible effect on the analysis.

- Removing all variables with 1 unique value.

- Standardizing date columns to Date Type.

- Removing all the variable with variance 0 for integer variable.

- Selected only numeric value of term and reassigned back to term.

- Selected only digit and < 1 does not extract, so for those blank assigned 0.

- Removing value 'Current' from the analysis.

Outlier Analysis

- 'annual_inc' having outliers.
- Removed 6 million salary which is outlier.
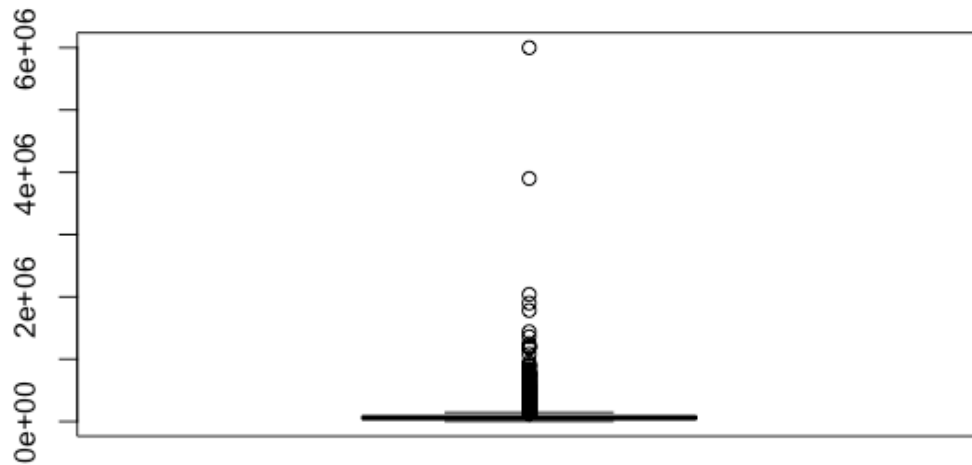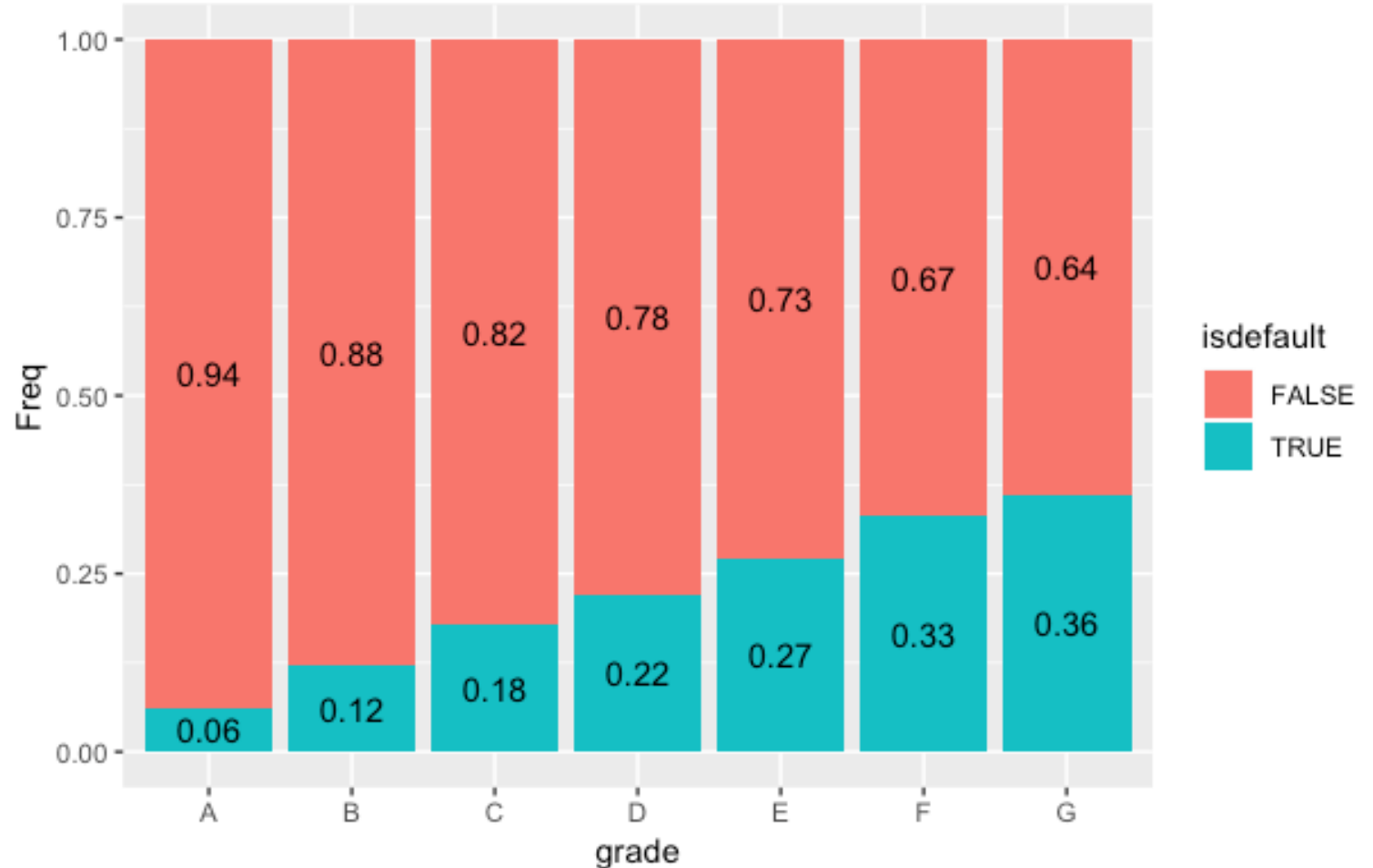

Before
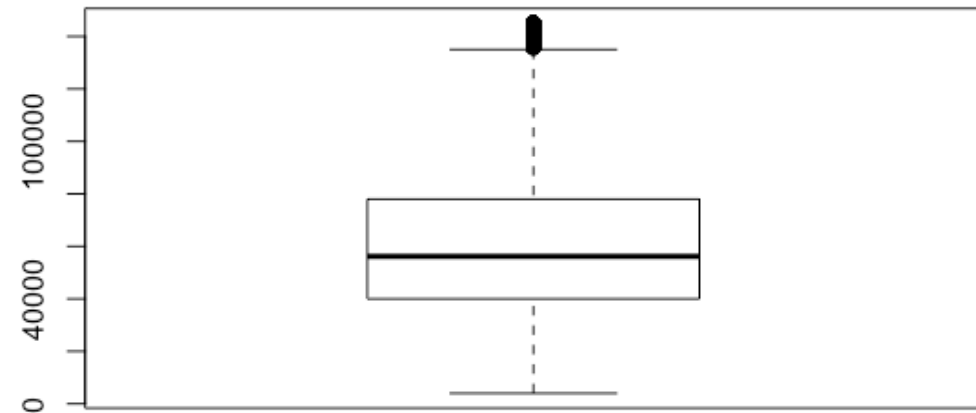
After

Segmented Univariate Analysis

Variable:

 Grade

Impact:

As grade go from A-G
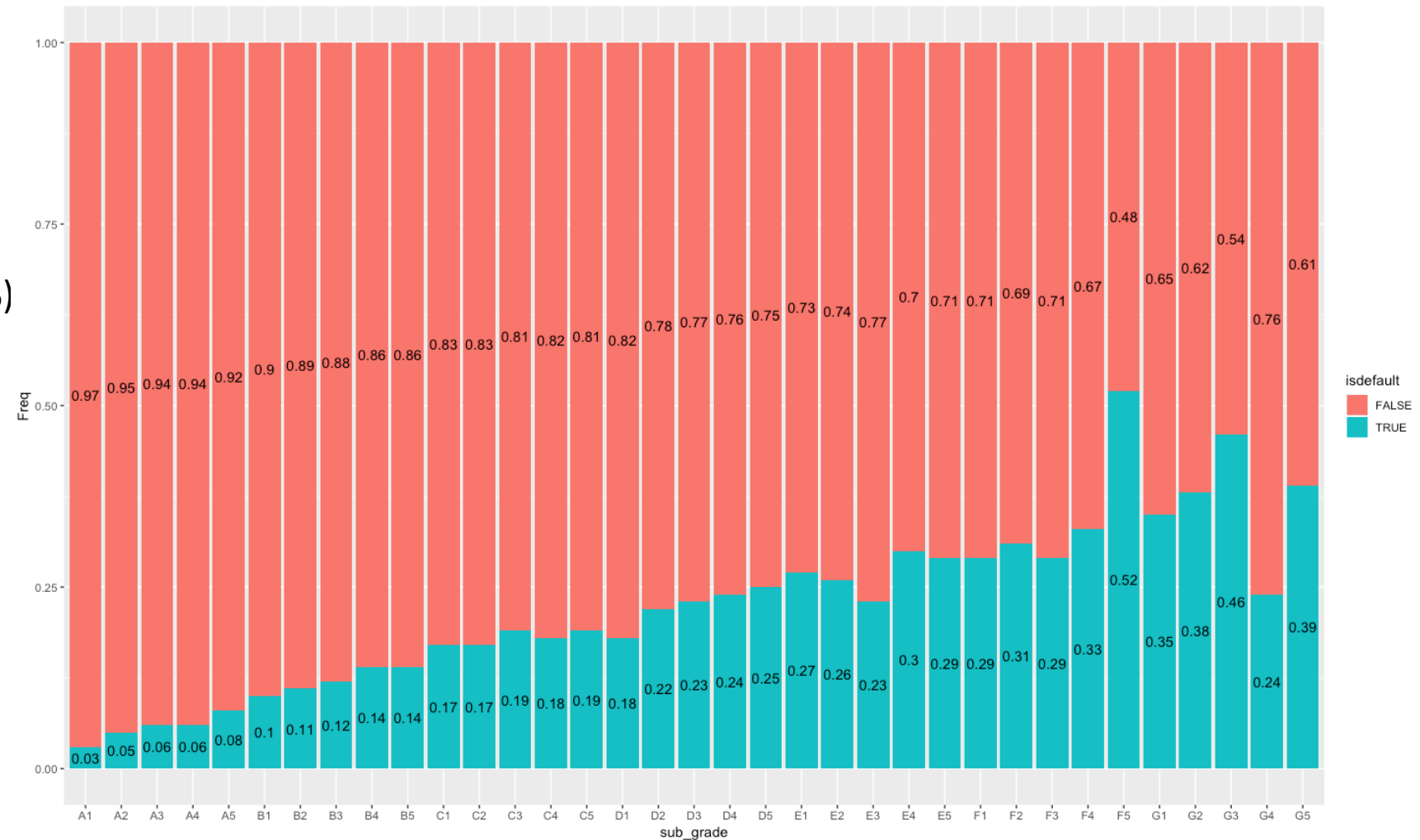chances of defaulting is
increasing

Segmented Univariate Analysis

Variable:

 Sub - Grade

Impact:

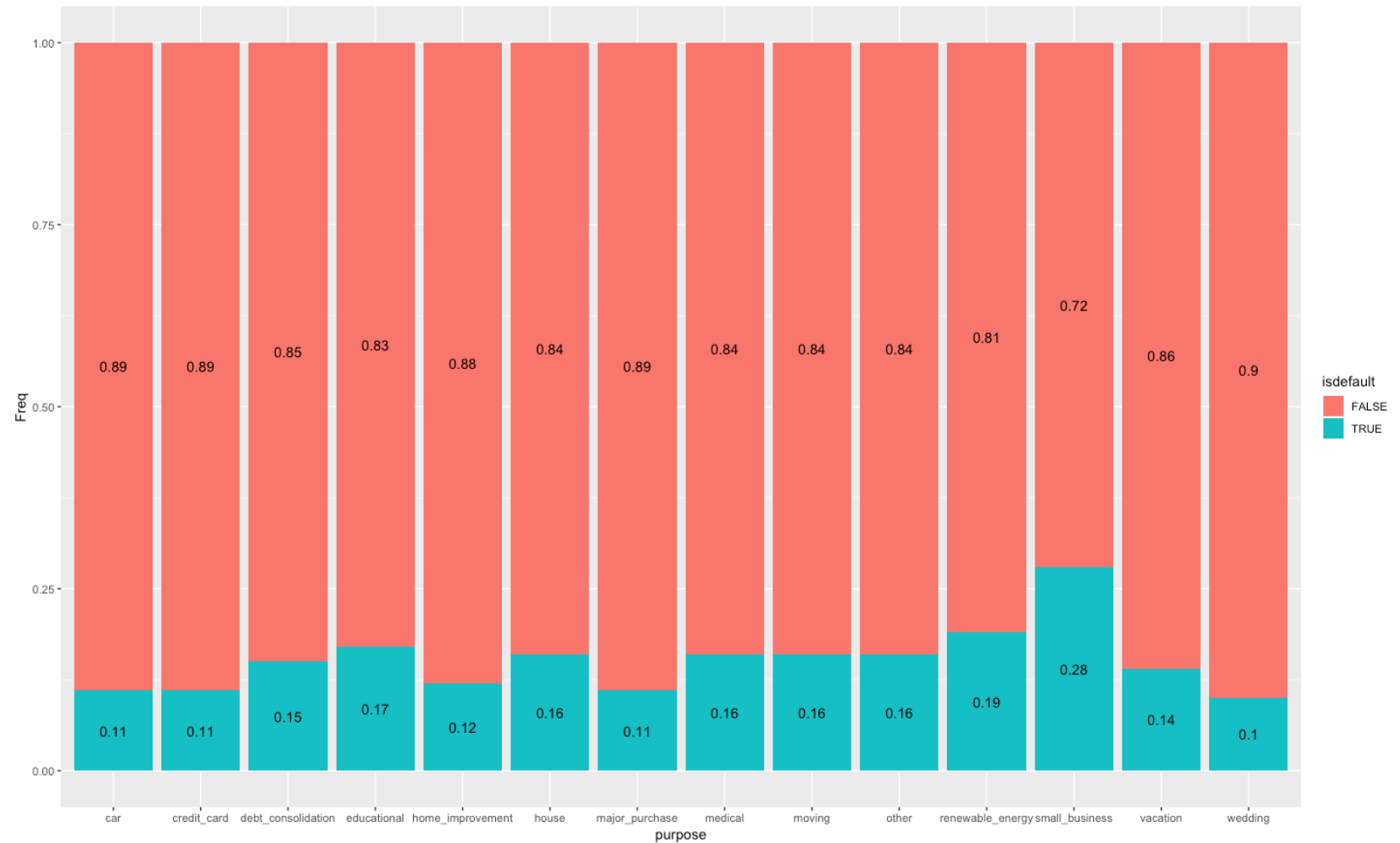F5 and G3 have more chance of defaulting (40%)

Segmented Univariate Analysis

Variable:

Purpose

Impact:

Small Business are having more number of defaults around (28%)
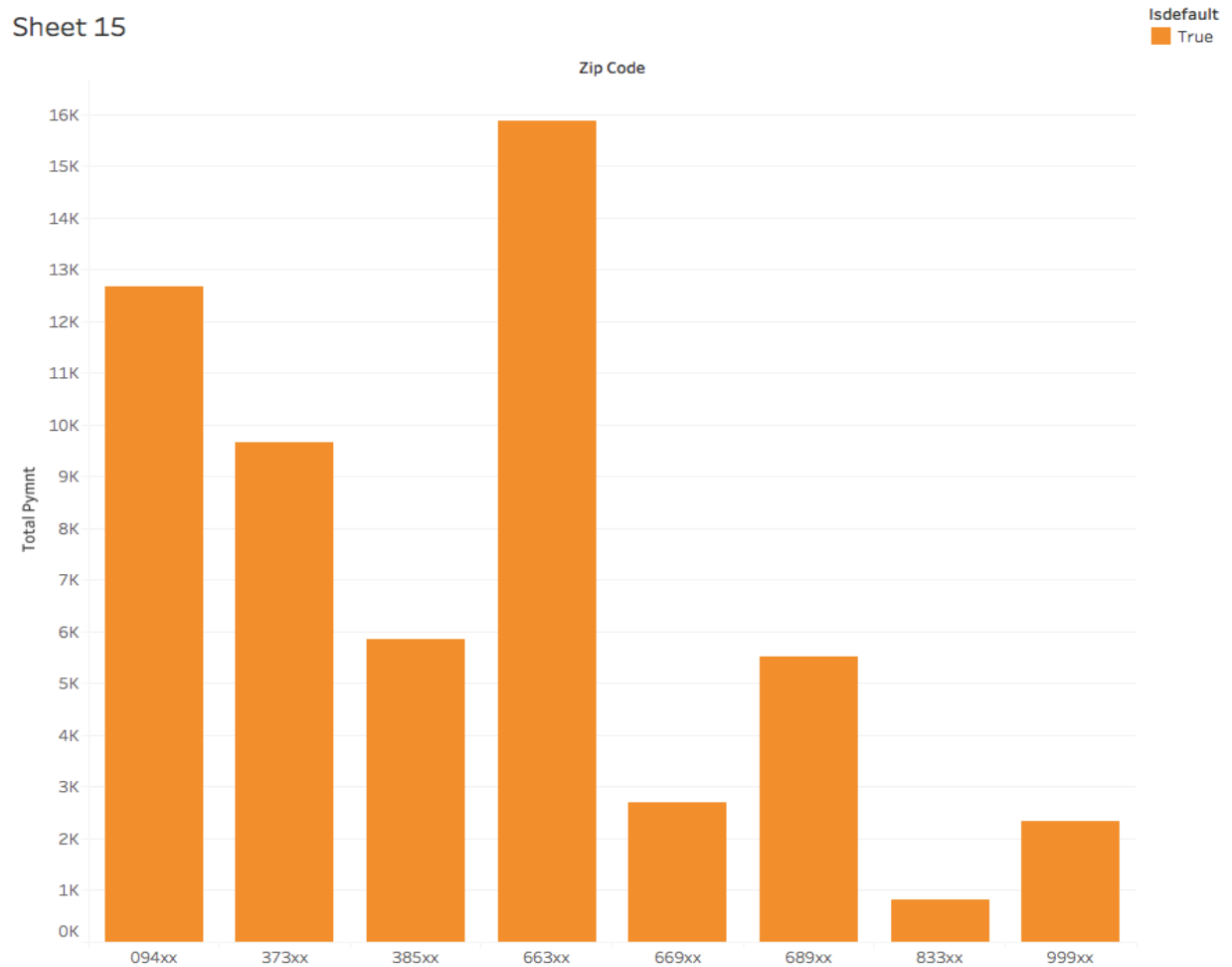
Segmented Univariate Analysis

Variable:

Zip code
(Plotting only those zipcodes with high number of defaults as the number of unique values is high)

Impact:

In these zip codes more than 90% cases are defaulting. 094xx 373xx 385xx 663xx 669xx 689xx 833xx 999xx
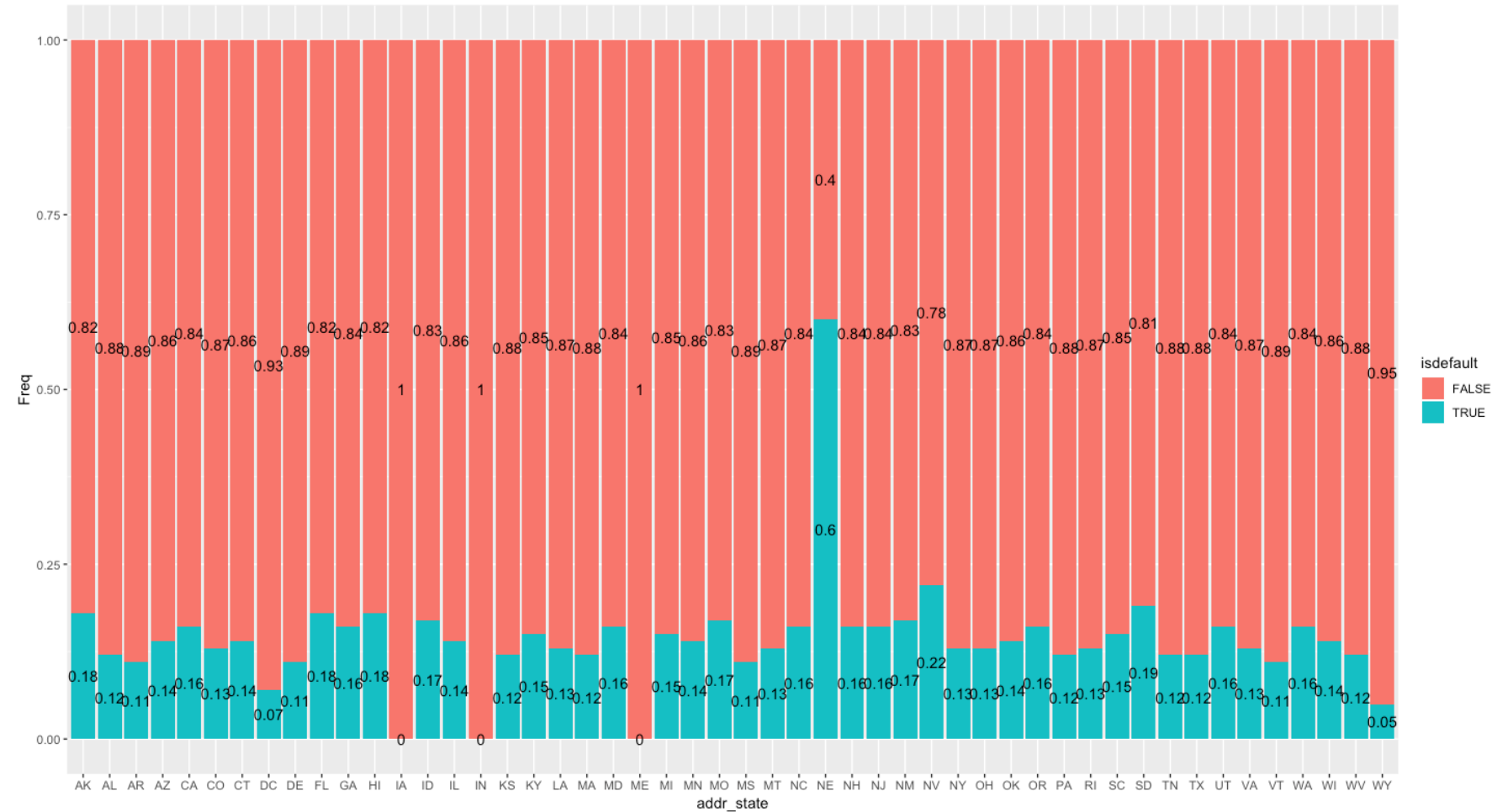
Segmented Univariate Analysis

Variable:

 State

Impact:

Nebraska (NE) having highest ratio of defaulters,
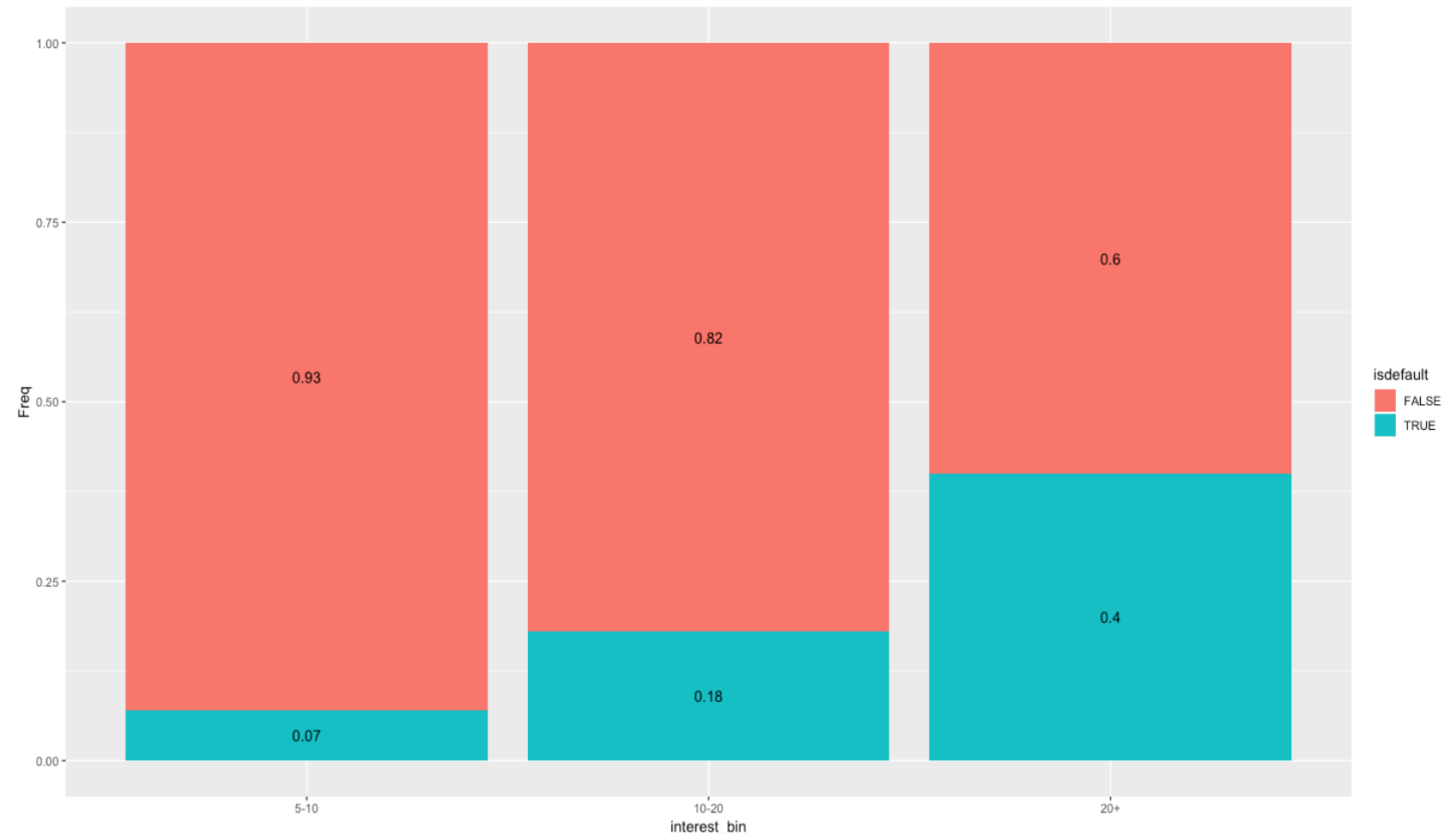But the number of records is very less.

Segmented Univariate Analysis

Variable:

 Interest Rate

Impact:

Number of defaulters is
more for interest rate
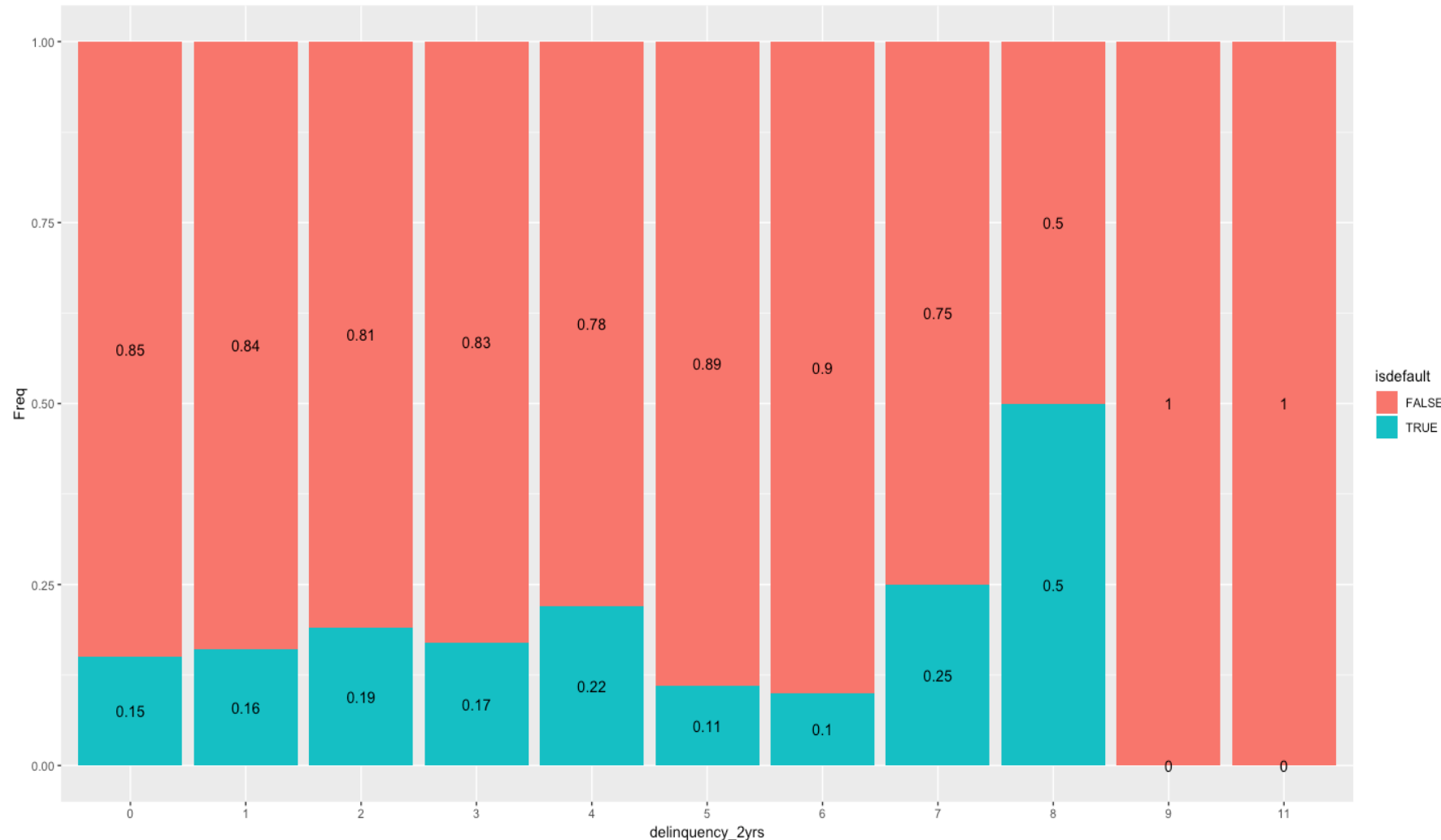20+ .

Segmented Univariate Analysis

Variable:

delinquency_2yrs

Impact:

If candidate had delinquent 8 times then around 50% are defaulting. Surprisingly delinquent more than 9 to 11 does not show any defaults. It may be because of data issue or insufficient data.
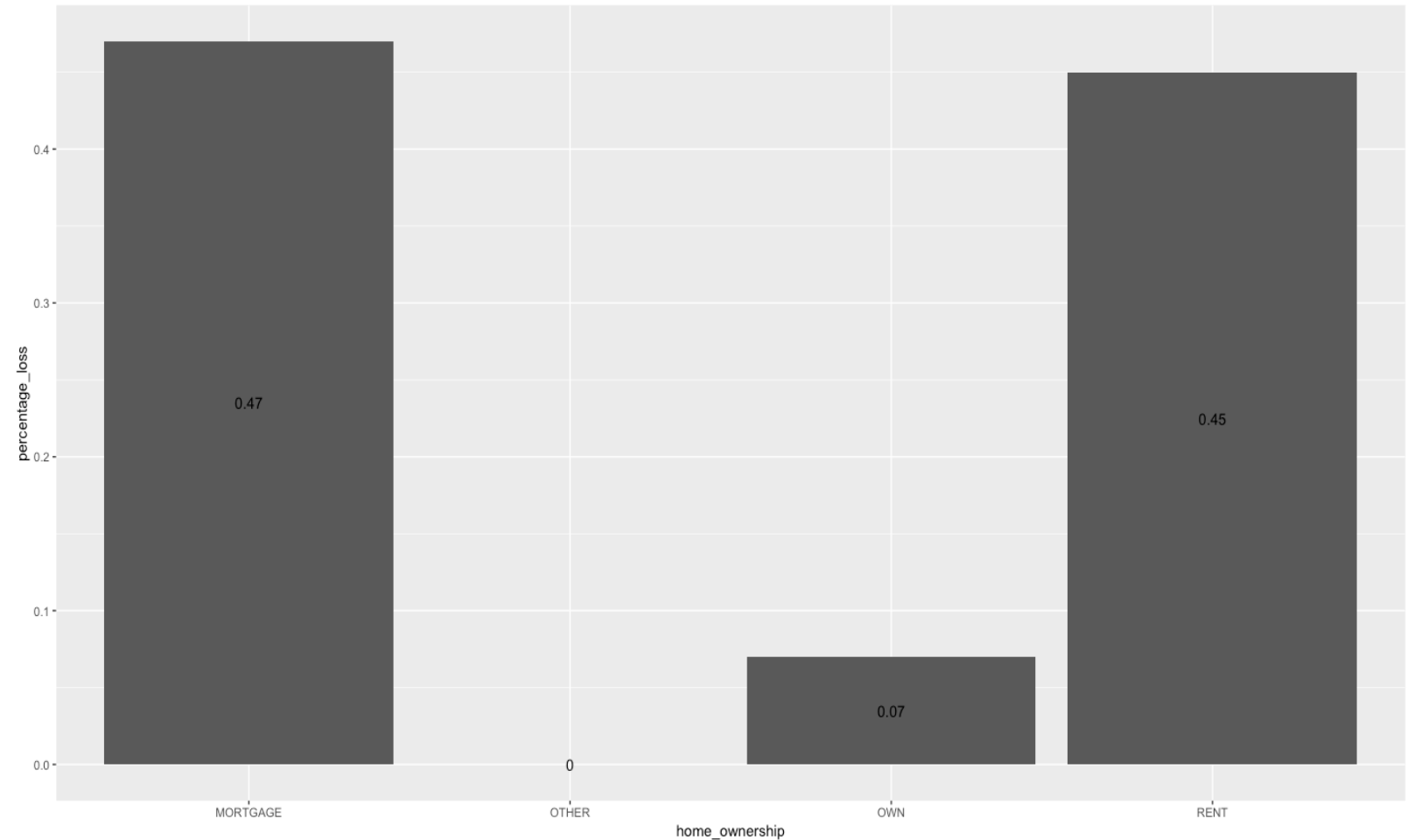
Univariate Analysis

Variable:

home_ownership vs loss

Impact:

User having MORTGAGE property tendency to defaulting is high 47% and even for rented which is around 45%
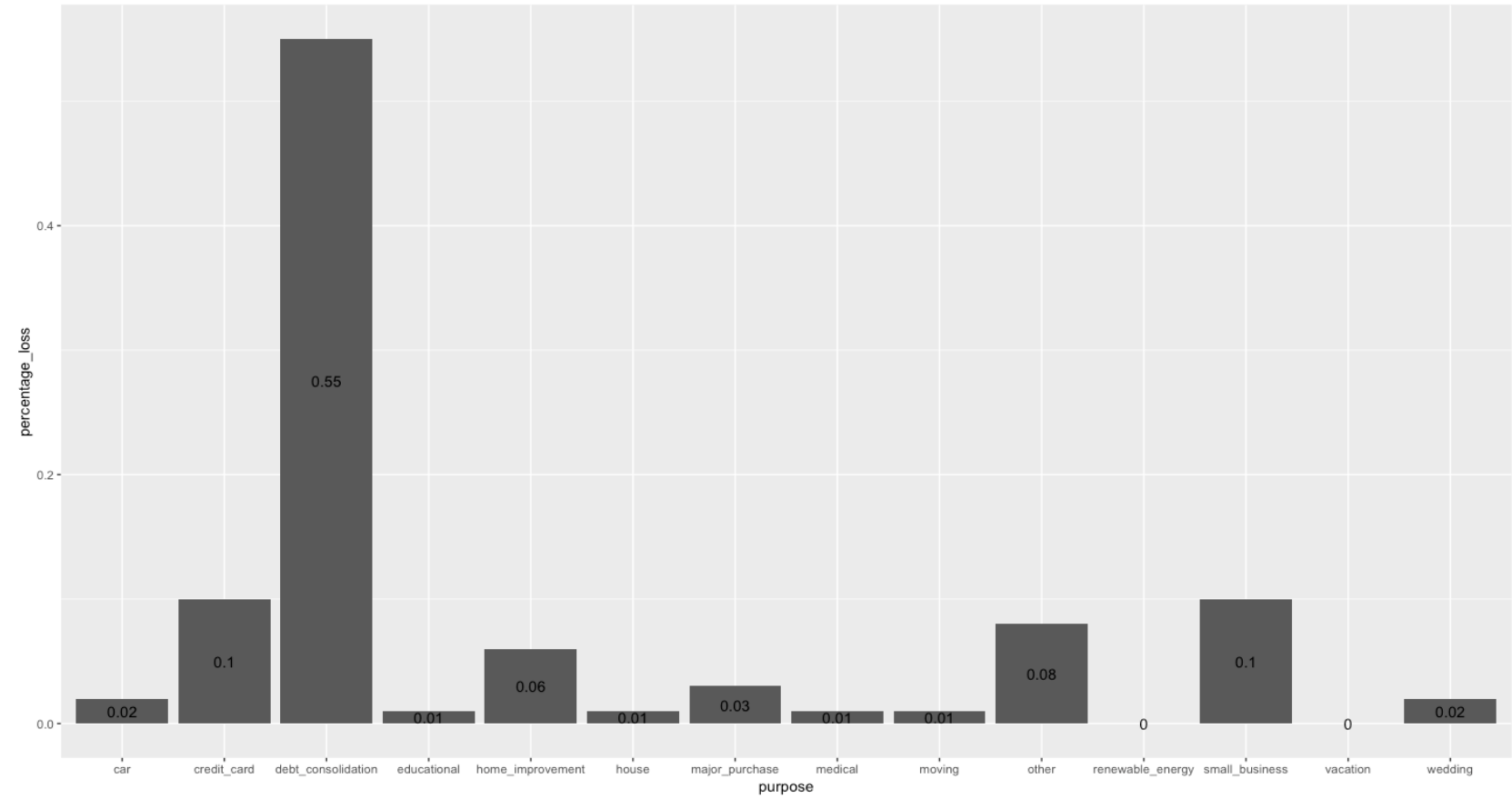
Univariate Analysis

Variable:

Purpose vs loss

Impact:

debit_consolidation (55.5%) is contributing around 55% of loss, though number wise there is no sign but when it comes revenue this is major
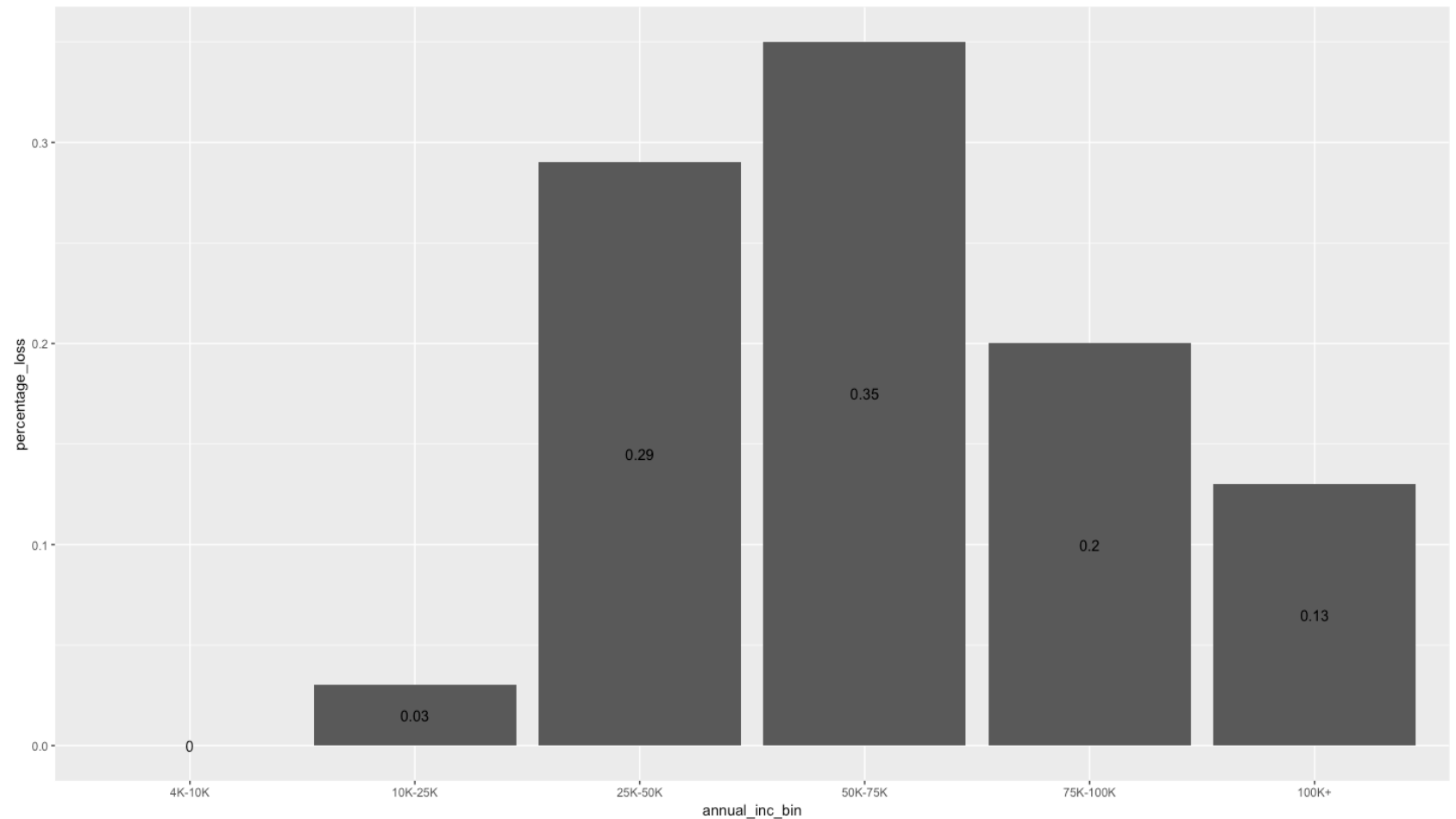
Univariate Analysis

Variable:

annual_inc vs loss

Impact:

Income having 50K-70K contributing 36% loss because of defaulting
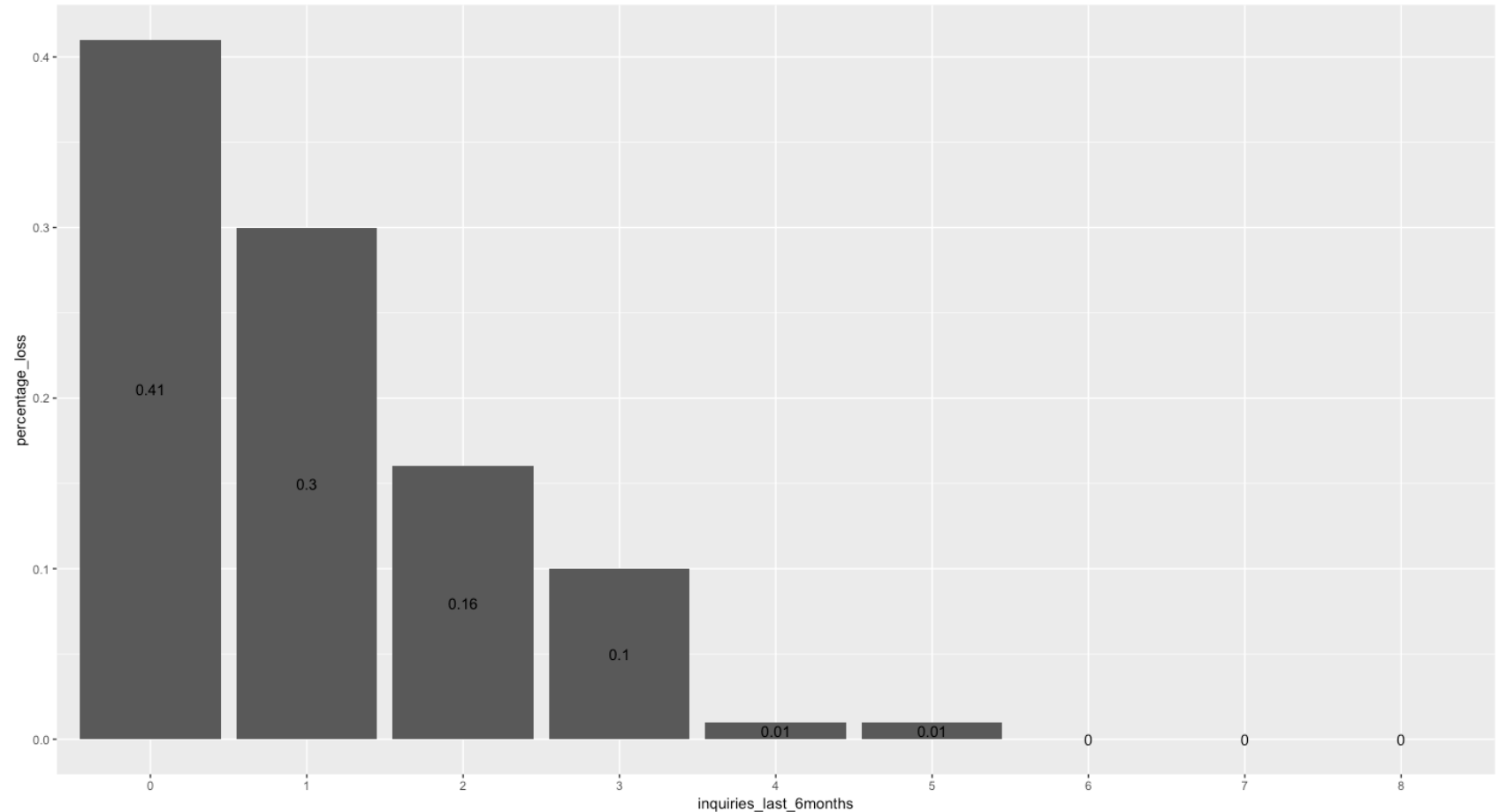
Univariate Analysis

Variable:

inquiries_last_6months vs loss

Impact:

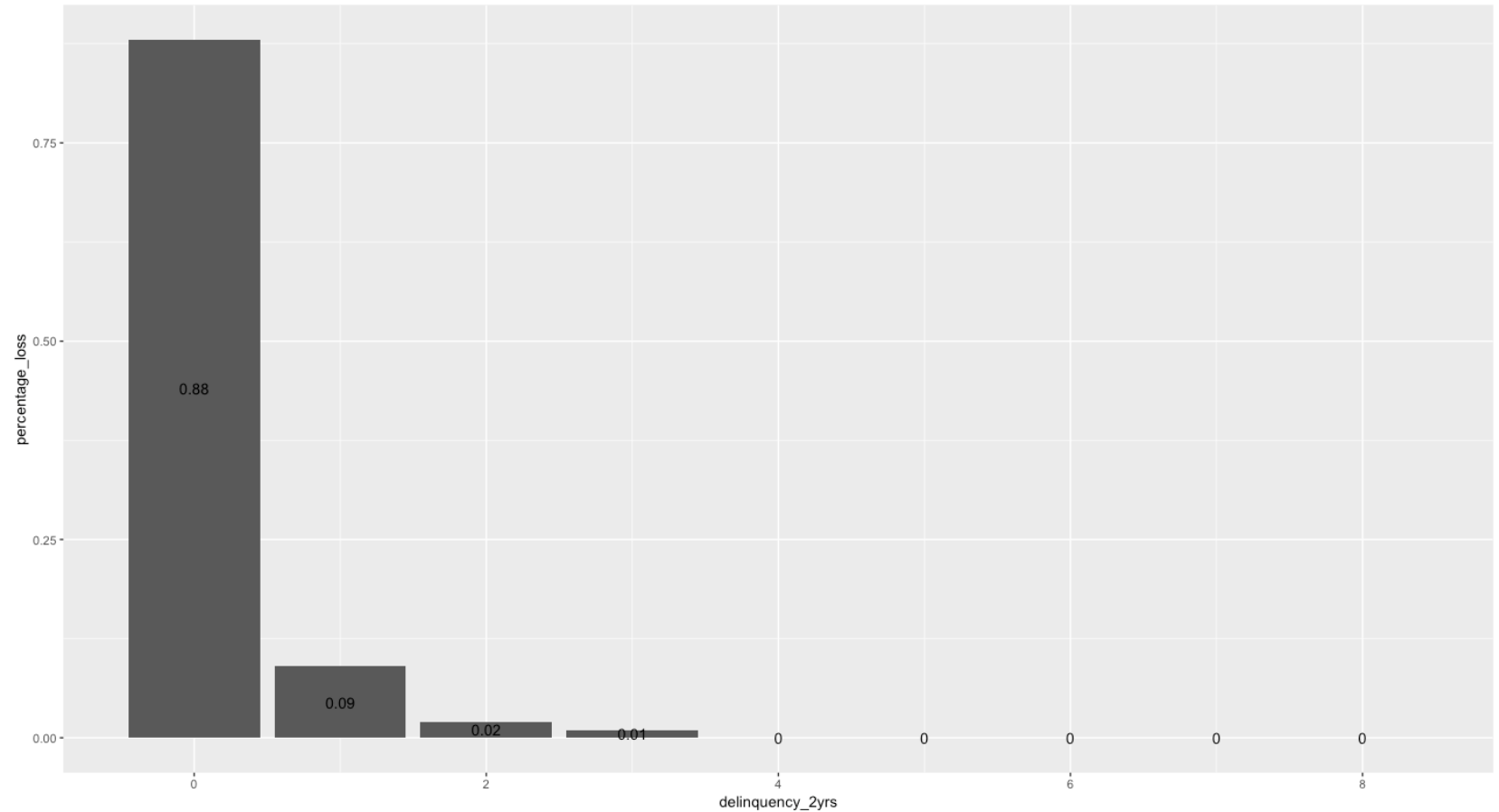Strange result no enquiries of credit line showing 40% loss

Univariate Analysis

Variable:

delinquency_2yrs vs loss

Impact:

Strange result without no delinquency showing 88% loss

Univariate Analysis
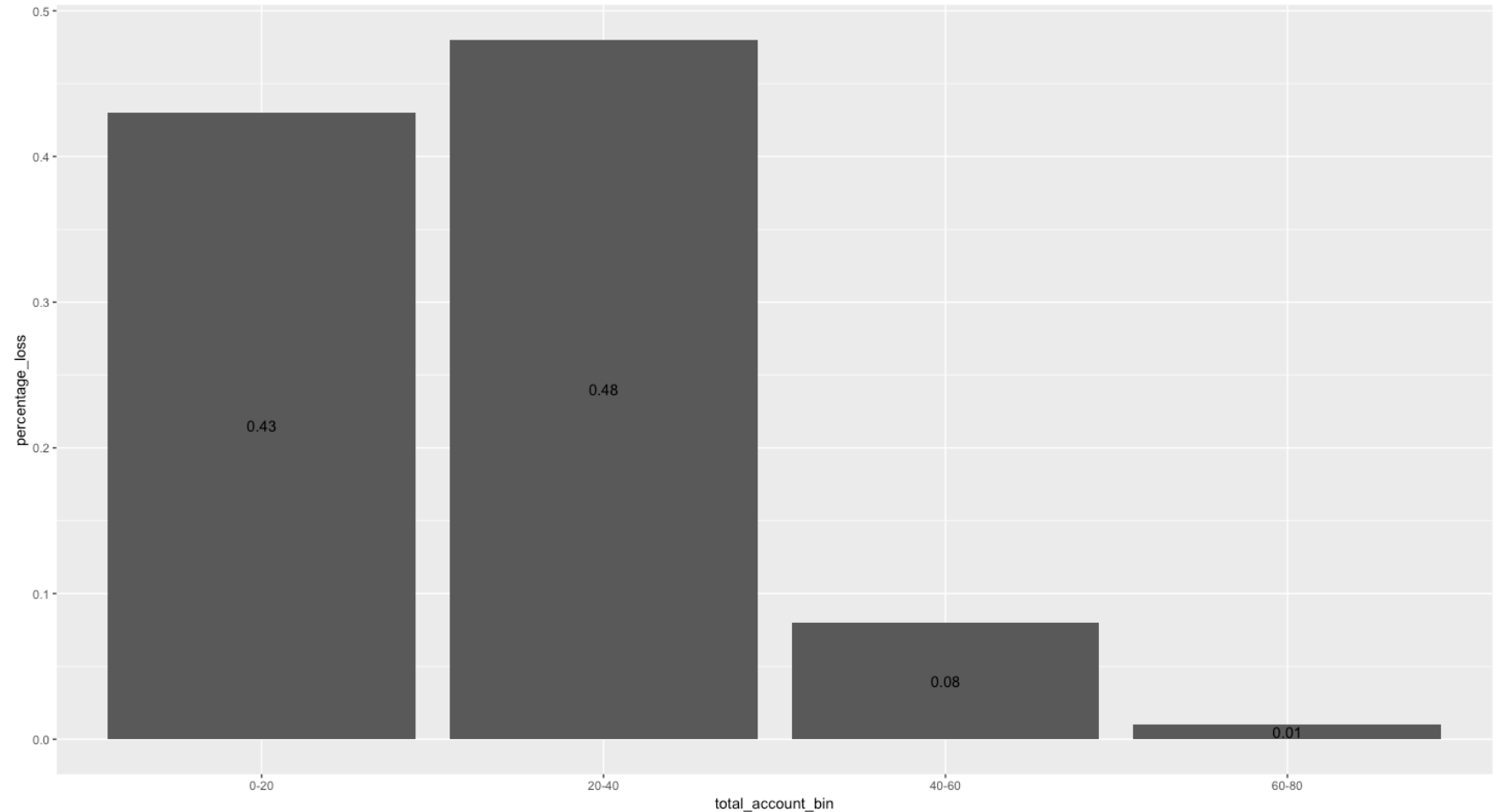
Variable:

total_account vs loss

Impact:

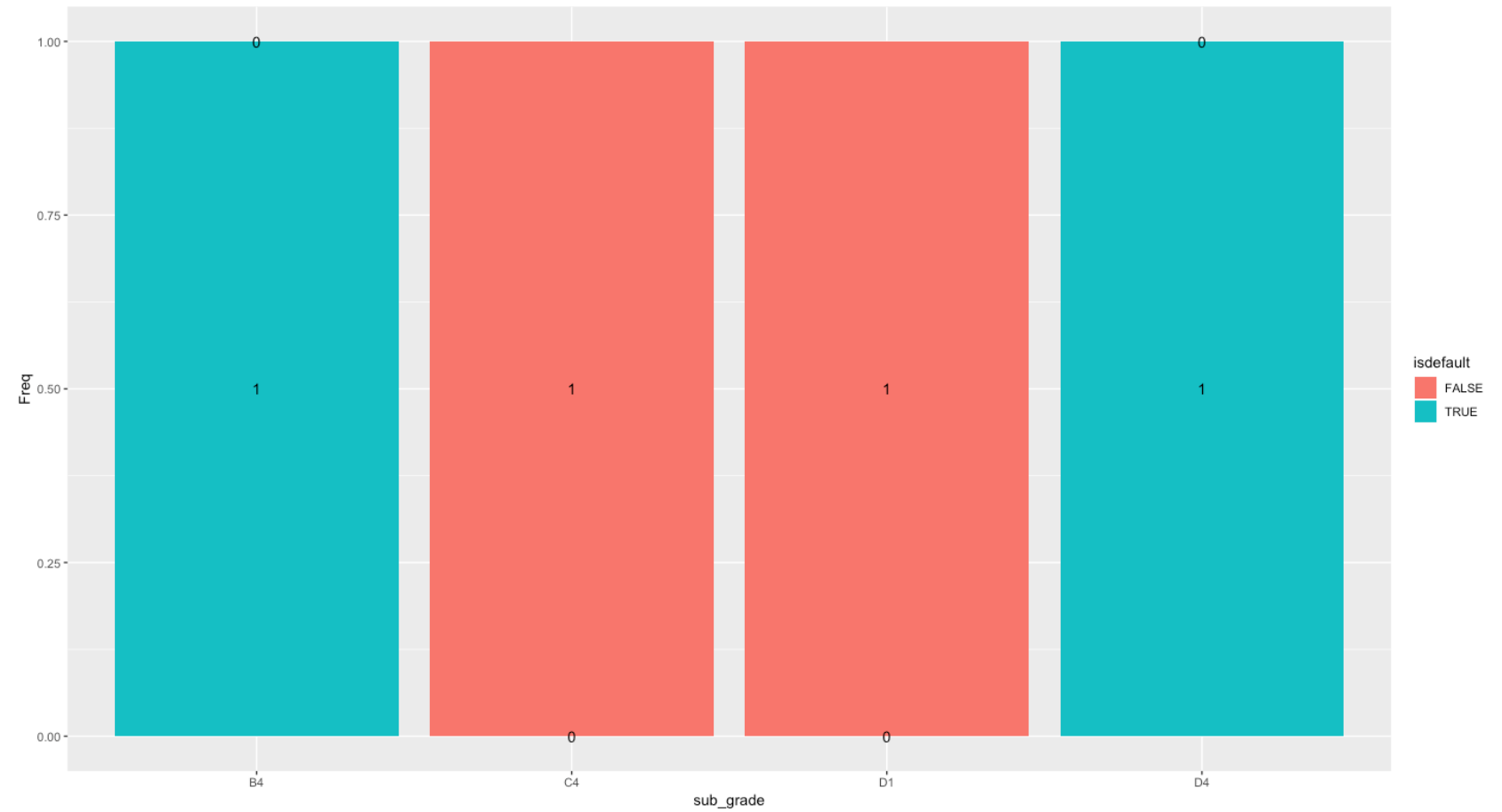If the current opened credit line <40 is causing major loss

Bivariate Analysis

Variable:

Sub grade for Nebraska (NE)

Impact:

Found that B4, D4 are defaulting more than 40%
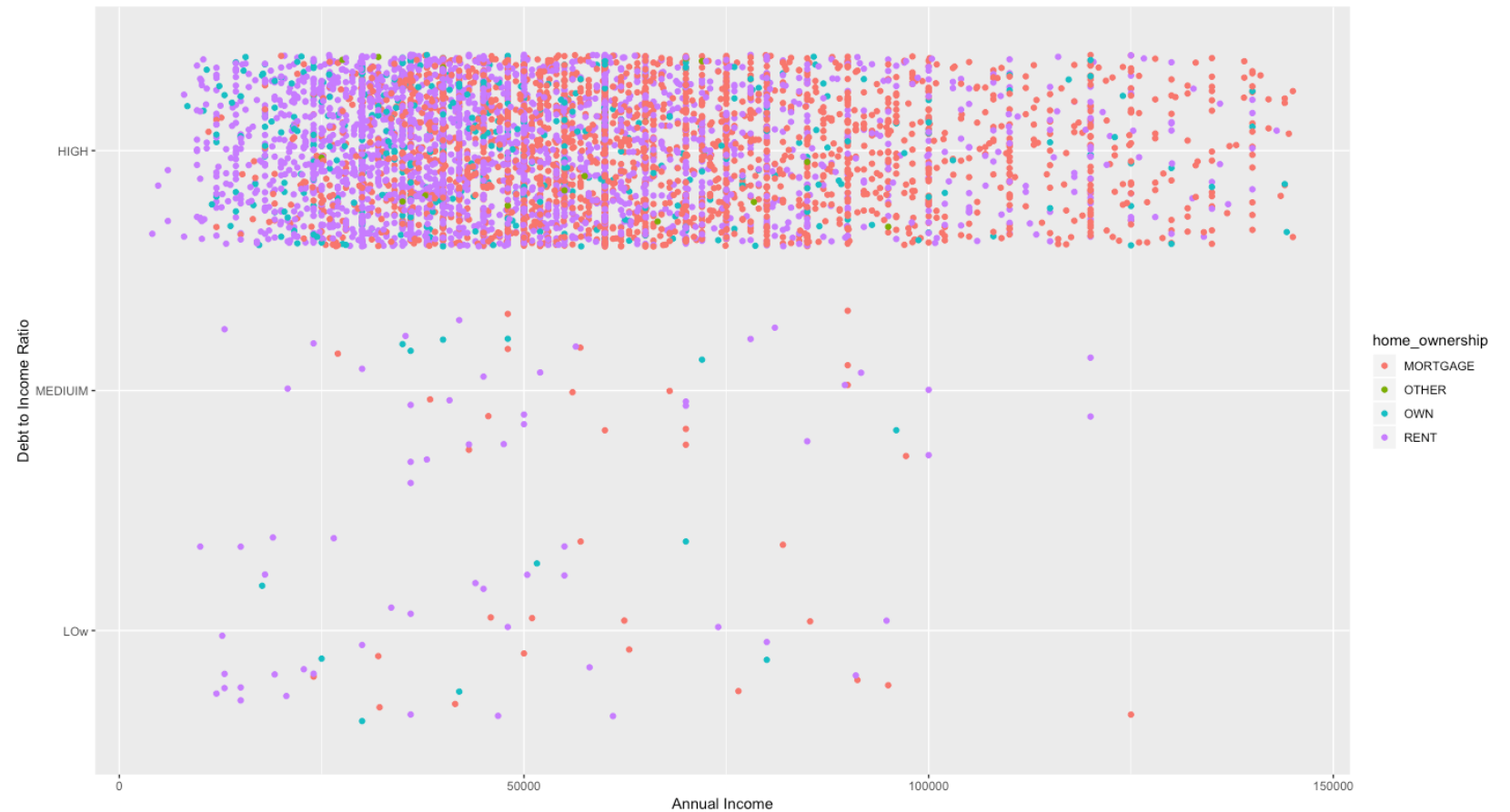
Bivariate Analysis

Variable:

Debt to Income and Home Ownership

Impact:

Applicant with annual income less than 50K and Dti ratio high (>1) and in Rented Home tend to default more.

Bivariate Analysis

Variable:

revolve_util and
revovle_bal

Impact:

Applicant with
revolving balance less
than 50K and revolving
utilization above 50%
tend to default more.
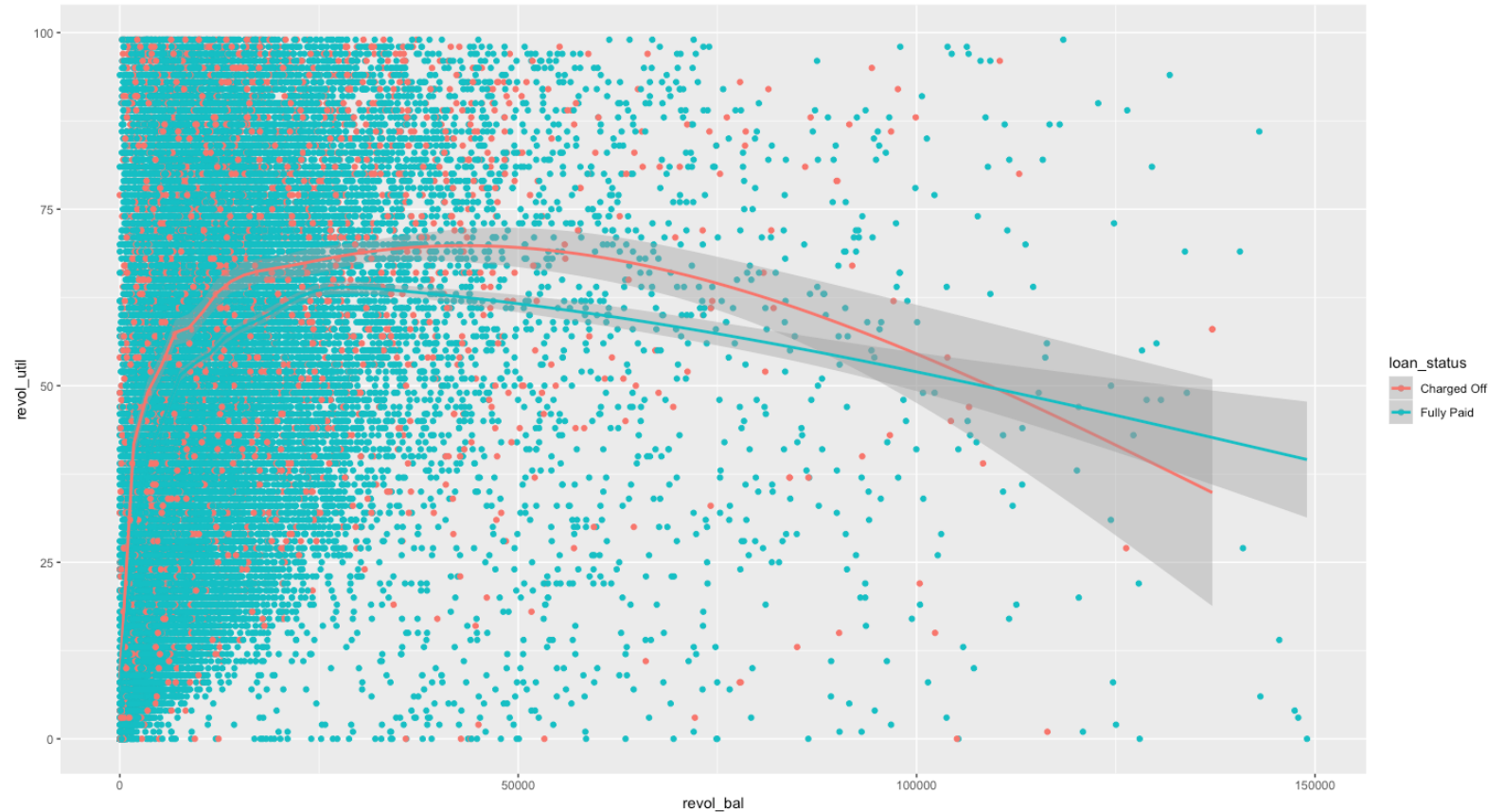
Correlation Analysis

- Installment has fairly strong correlation with loss
  0.9279056

- As loan amount increases loss to bank also increases because of default
  0.9998716

- Total Payment and loss are moderately correlated
  0.6771661

- Total rec int and loss are moderately correlated
  0.6930001

Analysis Conclusion

- Proportion of Loan defaulters are more with Higher Loan Amount
- Proportion of Loan defaulters increases with from LC Grades A to G
- No. of Applicants with Rented or with Mortgaged home are high , hence no. of defaulters among them are high
- Borrower's with Annual income less than 60K and with home ownership Rent and employee experience high are high defaulters
- Applicants with purpose of loan as debt consolidation , rented homes and Annual income 20K to 60K tend to default loans.
- Applicants with high Debt to income ration are risky.
- Percentage Applicants Nebraska (NE) State are maximum on defaulter's List.
- Applicants with Open credit line account with 2 are mostly to default.
- Applicants with revolving Credit Balance below 50K and Revolving line utilization rate with 50% and above tend to default loans.
- For interest rate above 20% the percentage of defaulters is the highest.

Recommendations

- Reduce funding amount for Applicants :
  - Purpose of Loan as Debt Consolidations
  - LC grade E,F,G
  - High Debt to Income Ratio i.e >1
  - Not Verified Source of Incomes

- Increase Interest Rates for Applicants with Not Verified Source of Income

Thank You!