

GRAMENER CASE STUDY

SUBMISSION

Group Name:

1. Shivam Kakkar
2. Sundeep Gupta
3. Nilaksh Bajpai
4. Rabbani Noorbash

Objective

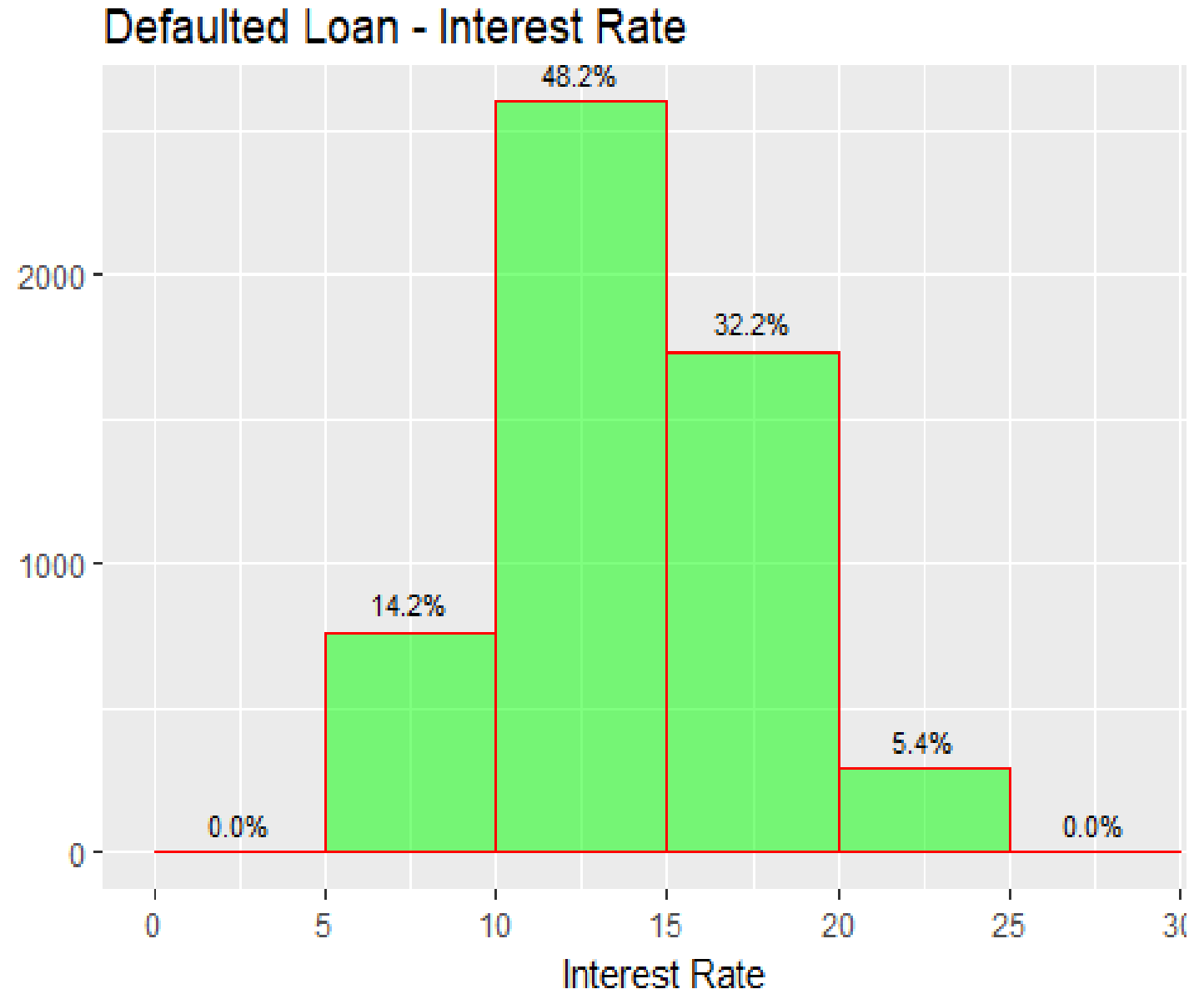
- This company is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures.
- To understand the factors due to which the loan is getting defaulted.
- Leverage the concepts of Exploratory Data Analytics such as Univariate / Bivariate Analysis of categorical and continuous variables to reach to a conclusion.

Univariate Analysis on Continuous variables

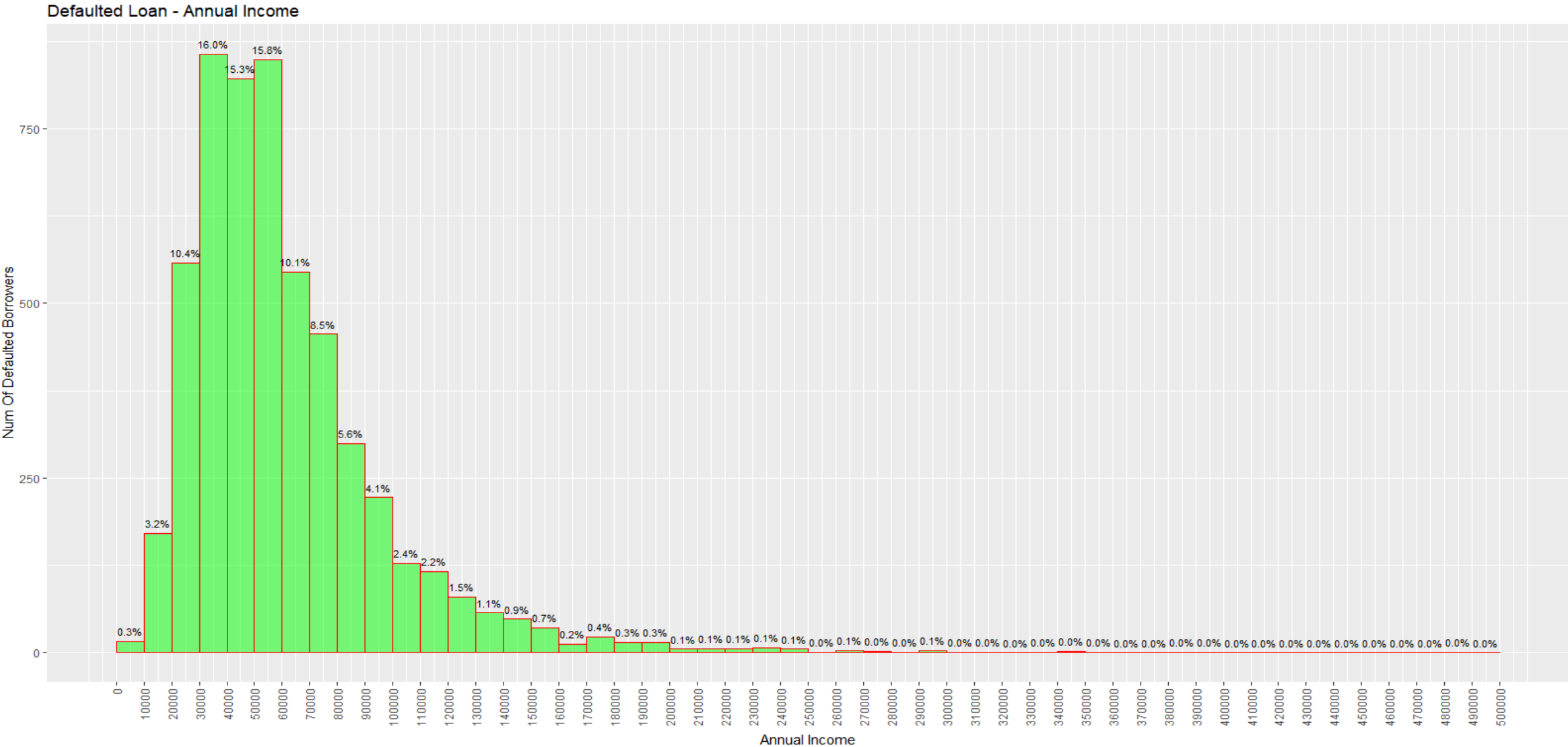
- Following variables are considered for Univariate Analysis
 - int_rate
 - funded_amnt
 - annual_inc
 - Installment
 - Debt to Income Ratio
 - Revolving Utilization
 - Revolving Balance
 - open_acc

int_rate

- A histogram for “int_rate” shows that majority of the default loans lies in the interest rate : 10 – 20%
 - 5 -10 % : **14.2** % of defaulted loans
 - 10 – 15 % : **48.2** % of defaulted loans
 - 15 – 20 % : **32.2** % of defaulted loans
 - 20 – 25 % : **5.4** % of defaulted loans
- From this we have created a derived metric : “int_range” as a categorical variable:
 - 0 - 10
 - 10 - 20
 - > 20



Annual Income

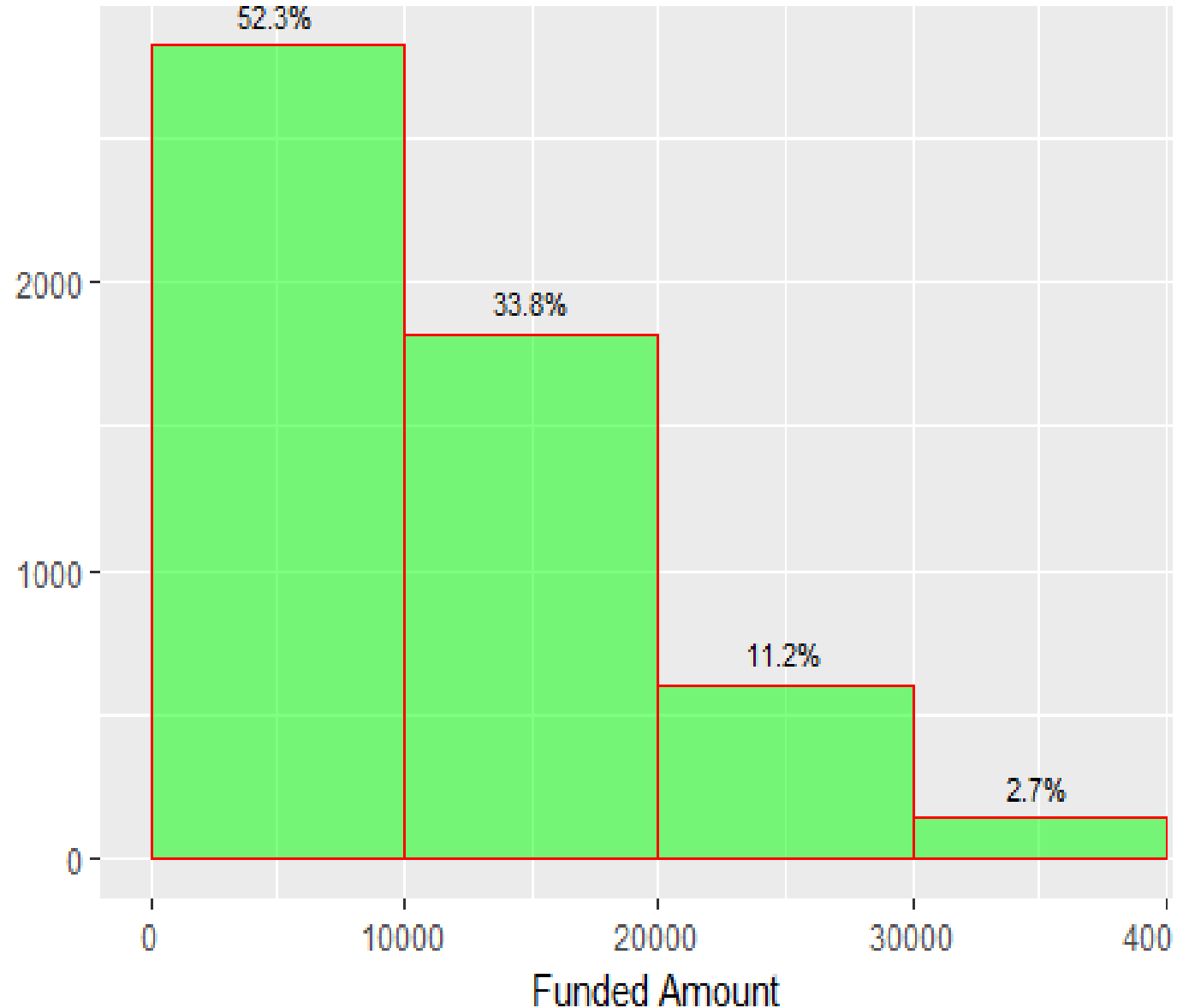


- From the above “annual income” histogram it is clearly visible:
 - 0 - 20 K = **3.5 %**
 - 20K - 80K = **75.9%**
 - > 80K = **20.6 %**
- **75.9%** of the borrowers defaulted are earning between **30K – 80K**
- Created a new derived metric : “**annual_income_range**”:
 - LOW : 0 -20K
 - MEDIUM : 20K – 80K
 - HIGH : > 80K

Funded_amnt

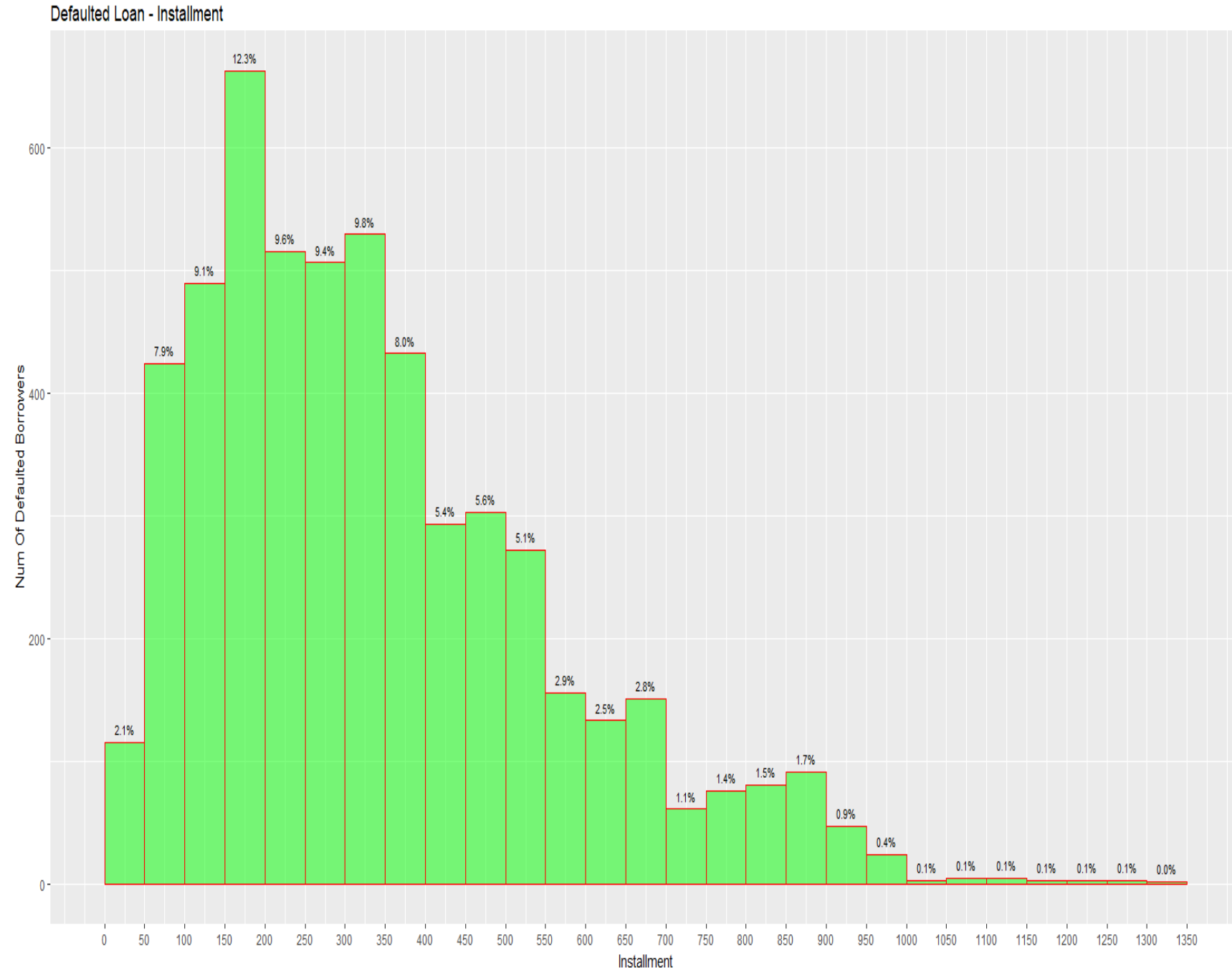
- From this it is clearly visible that majority of the defaulted loan lies in the range of : 0 -10K – **52.3%**
 - 0 – 10K : 52.3 % of borrowers defaulted.
 - 10K – 20K : 33.8 % of borrowers defaulted.
 - 20K – 30K : 11.2 % of borrowers defaulted.
 - > 30K : 2.7 % of borrowers defaulted
- From this we have created a derived metric : “funded_amnt_range” as a categorical variable:
 - LOW : 0 – 10K
 - MEDIUM : 10K – 20K
 - HIGH : 20K – 30K
 - VERY HIGH : > 30K

Defaulted Loan - Funded Amount



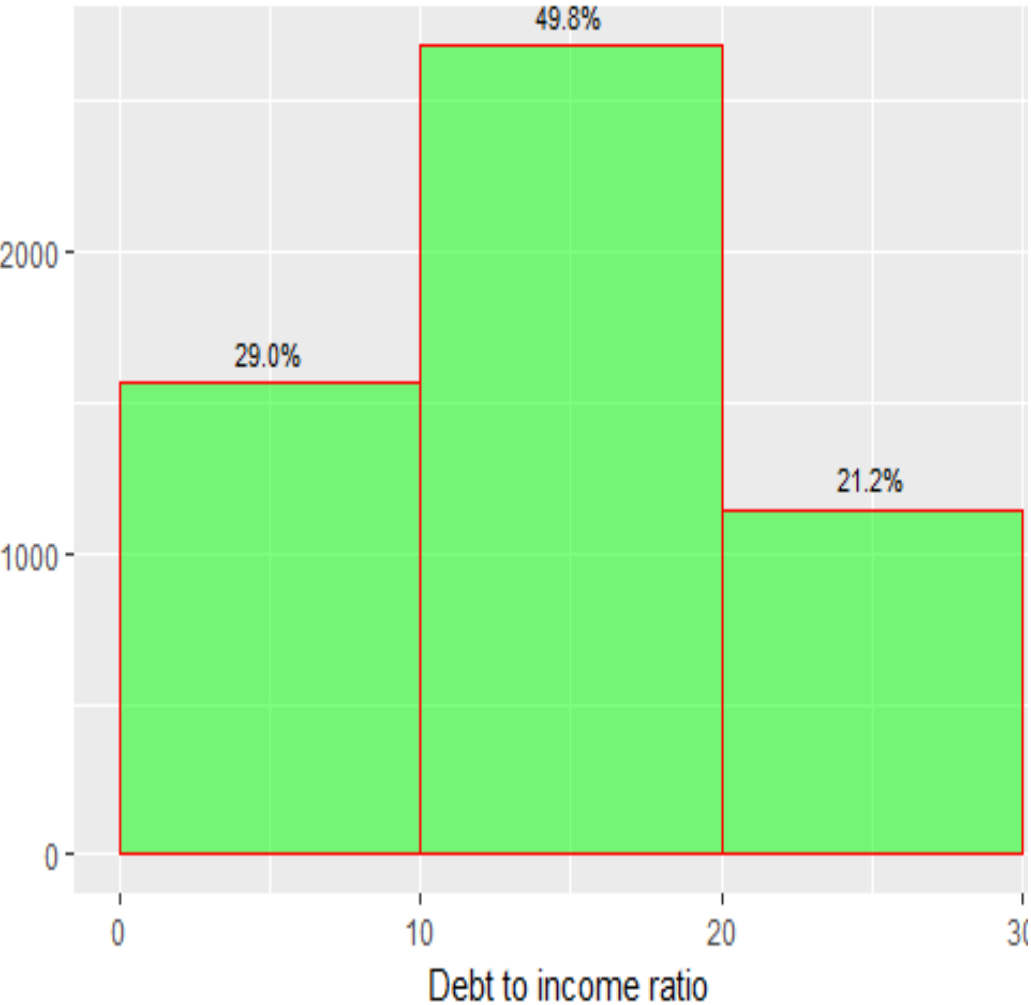
Installment

- A histogram for “installment” shows that majority of the default loans lies in : 150 -400 bracket.
 - 0 - 150 : **19.1%** of defaulted loans
 - 150 -400 : **49.1%** of defaulted loans
 - > 400 : **31.8%** of defaulted loans
- From this we have created a derived metric : “installment_range” as a categorical variable:
 - 0 - 150 : “LOW”
 - 150 - 400 : “MEDIUM”
 - > 400 : “HIGH”

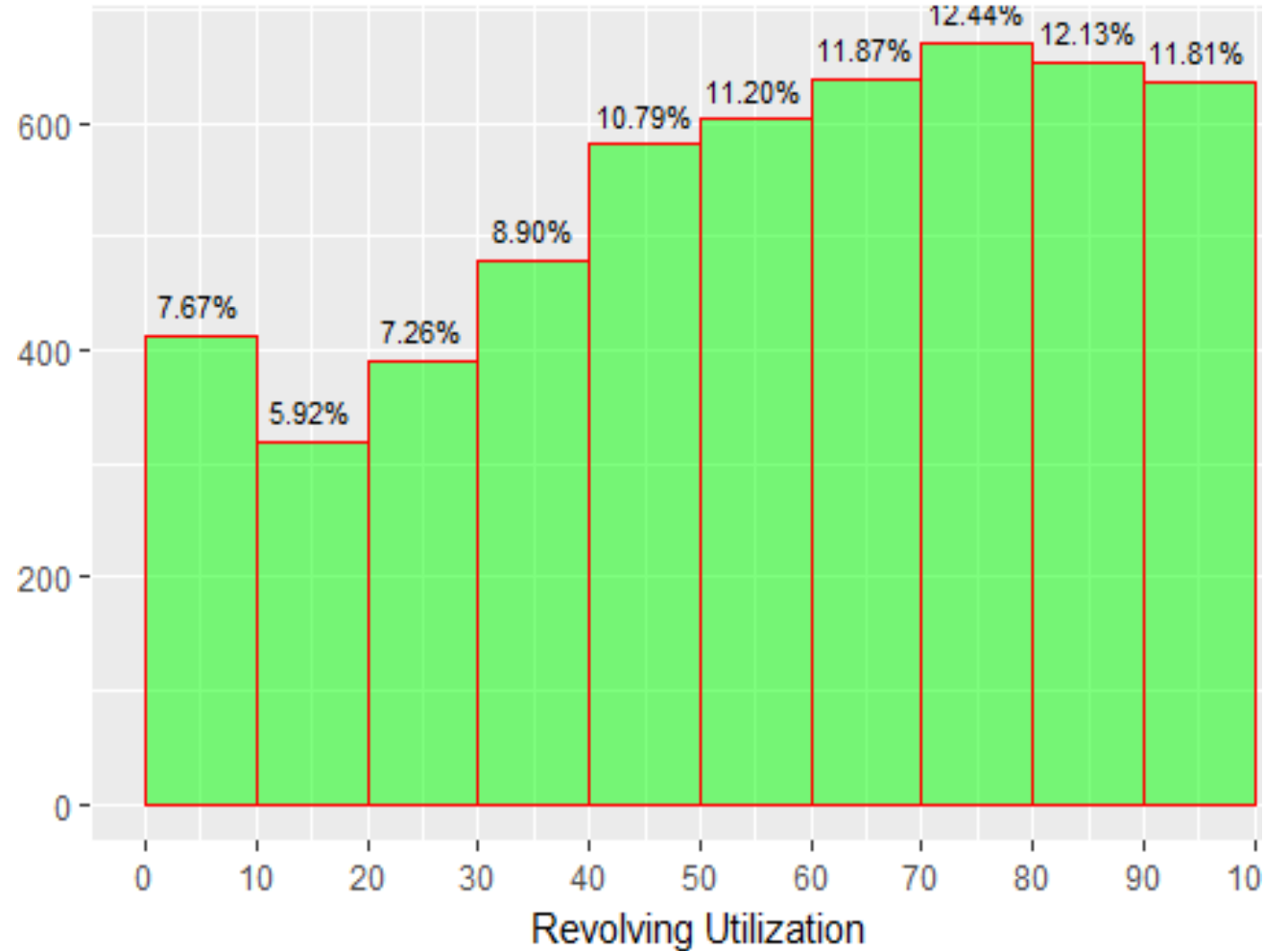


Debt to Income Ratio & Revolving Utilization

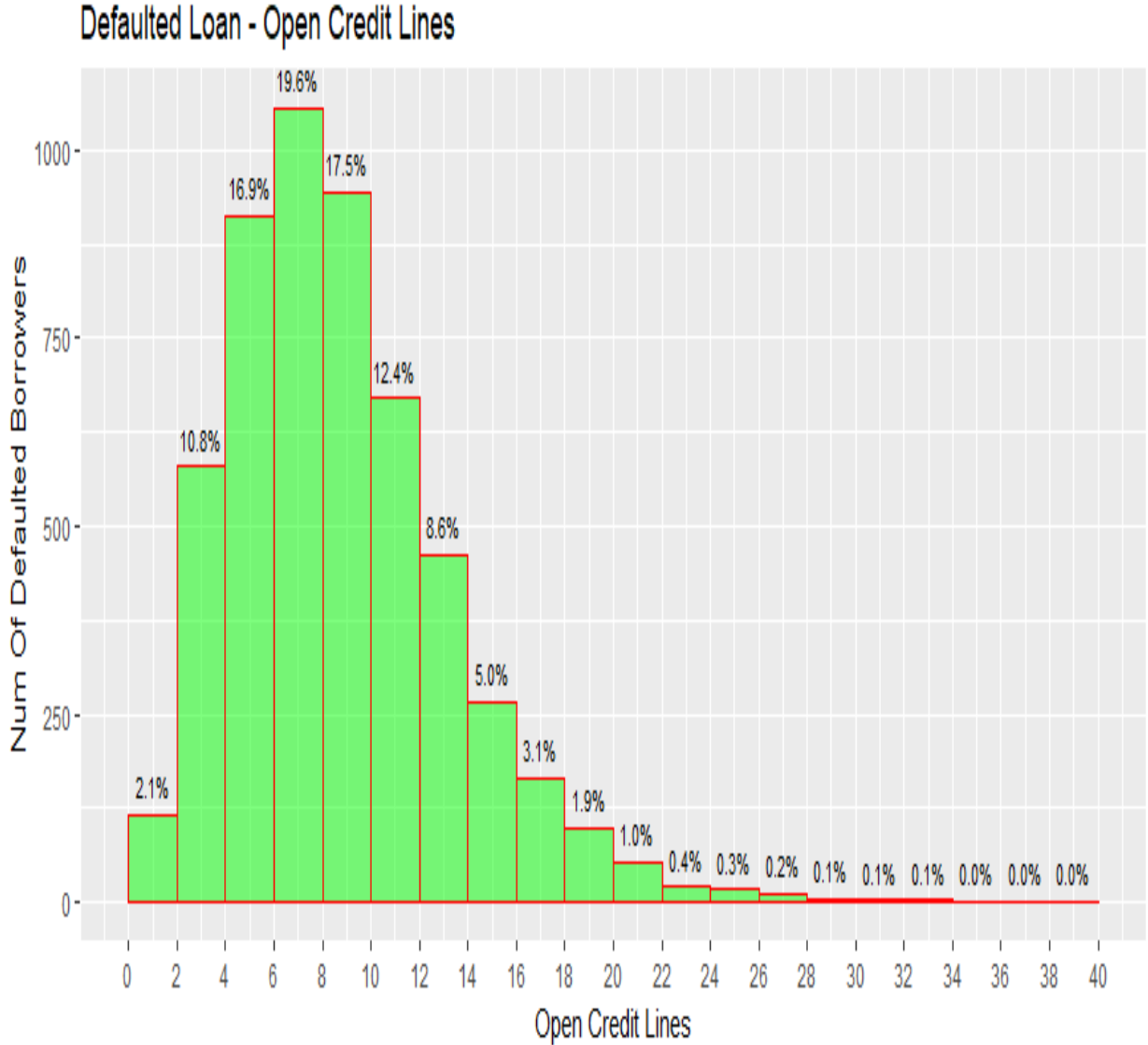
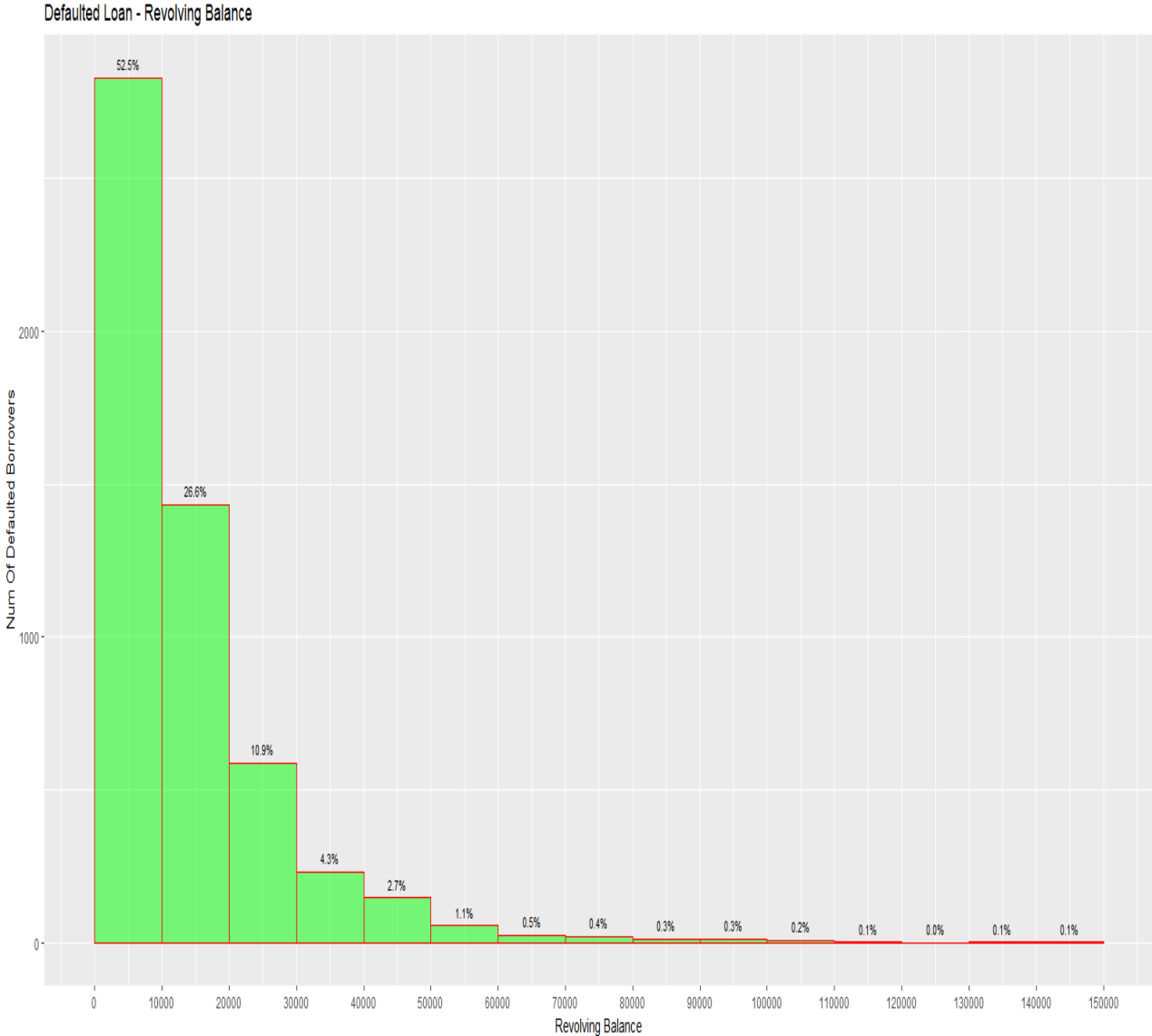
Defaulted Loan - DTI



Defaulted Loan - Revolving Utilization



For Revolving Balance and Open Credit Lines

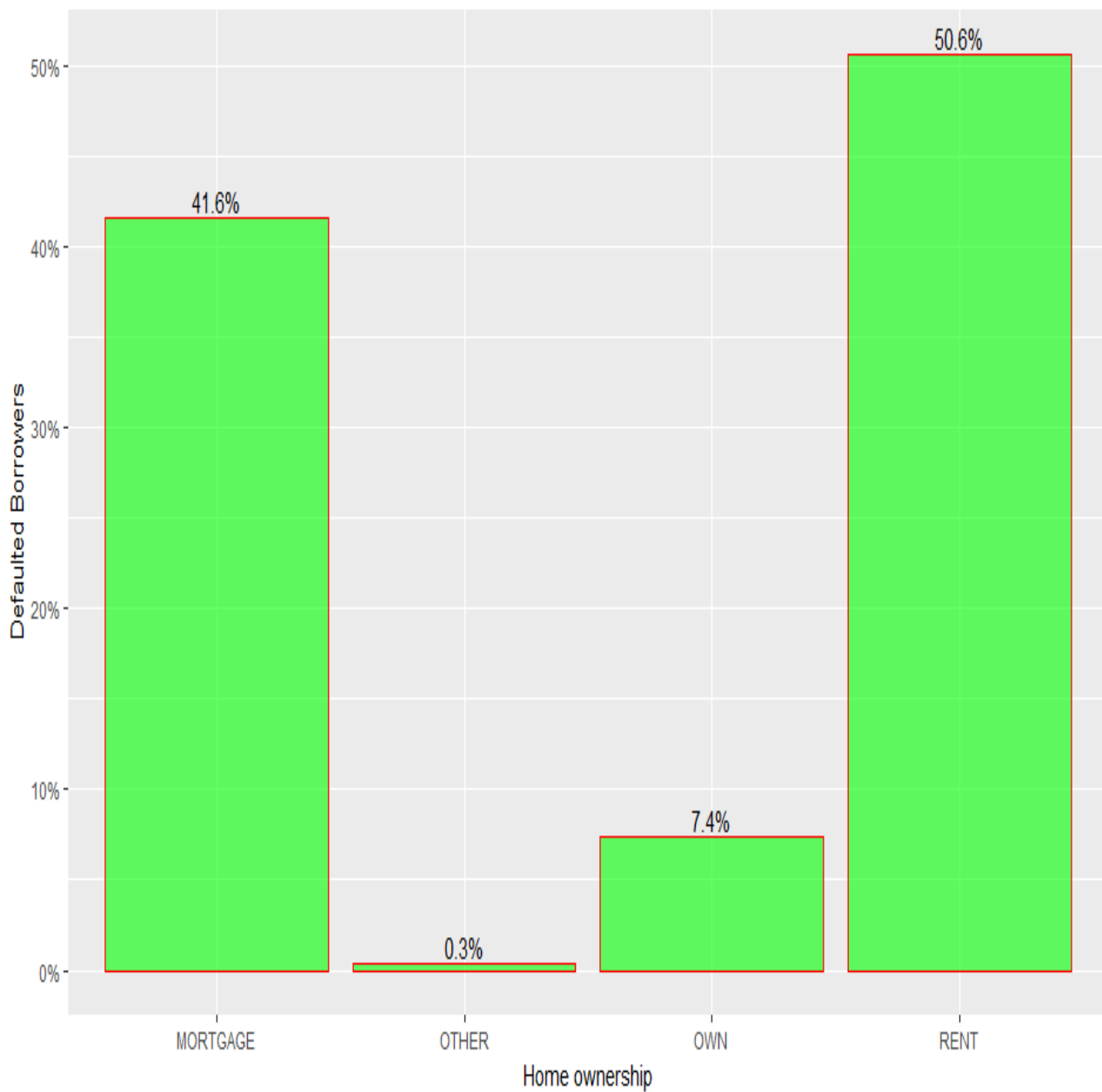


- From the above “open credit lines histogram” histogram it is clearly visible:
 - $0 - 4 = 2.1 \%$
 - $4 - 12 = 66.5 \%$
 - $> 12 = 31.4 \%$
- **66.5 %** of the borrowers defaulted are having credit lines between **4 -12**
- Created a new derived metric : “**open_acc_range**”:
 - $0 - 4$
 - $4 - 12$
 - > 12

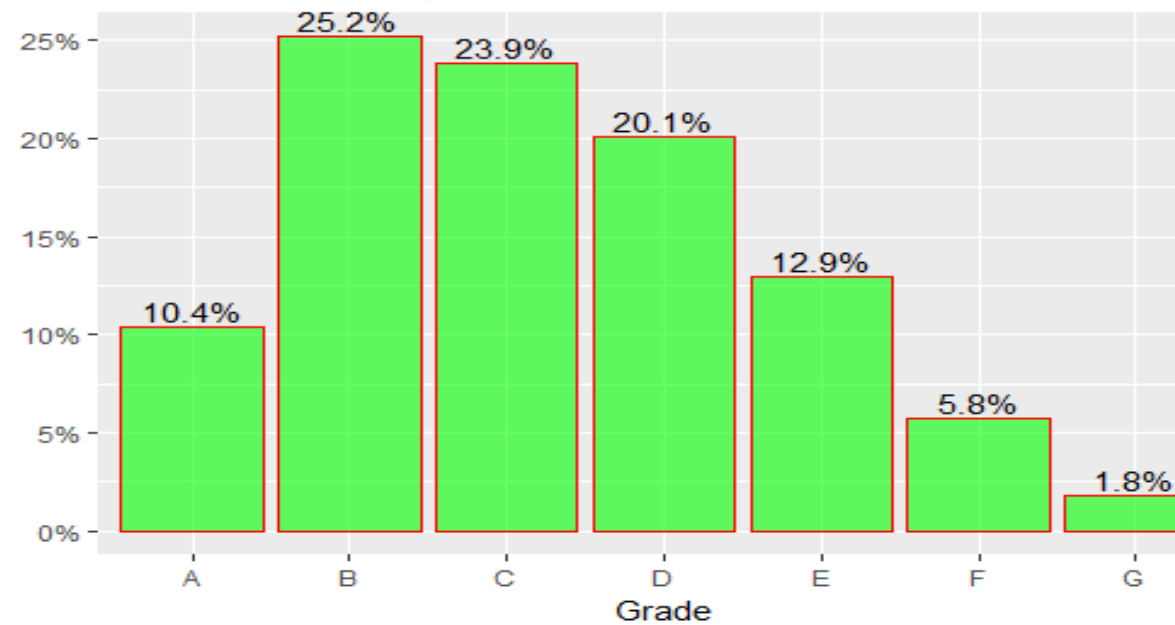
Univariate Analysis on Categorical variables

- Following variables are considered for Univariate Analysis
 - home_ownership
 - Verification_status
 - term
 - Purpose
 - Length of employment
 - Grade
 - Funded_amnt_range (derived metric)
 - Installment_range (derived metric)
 - Int_range (derived metric)
 - Annual_income_range (derived metric)

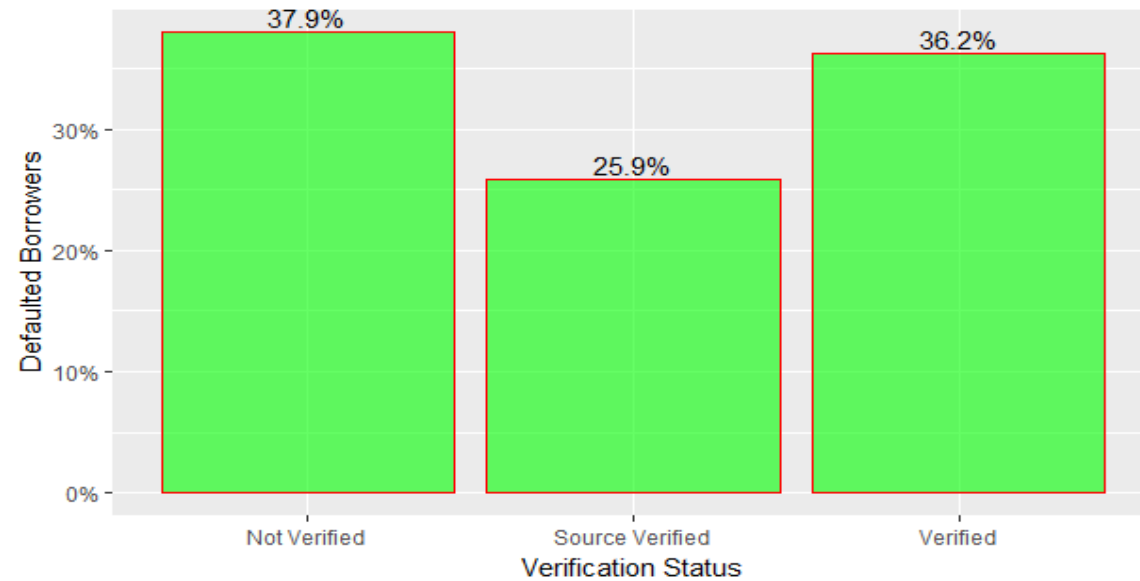
Defaulted loans - Home ownership



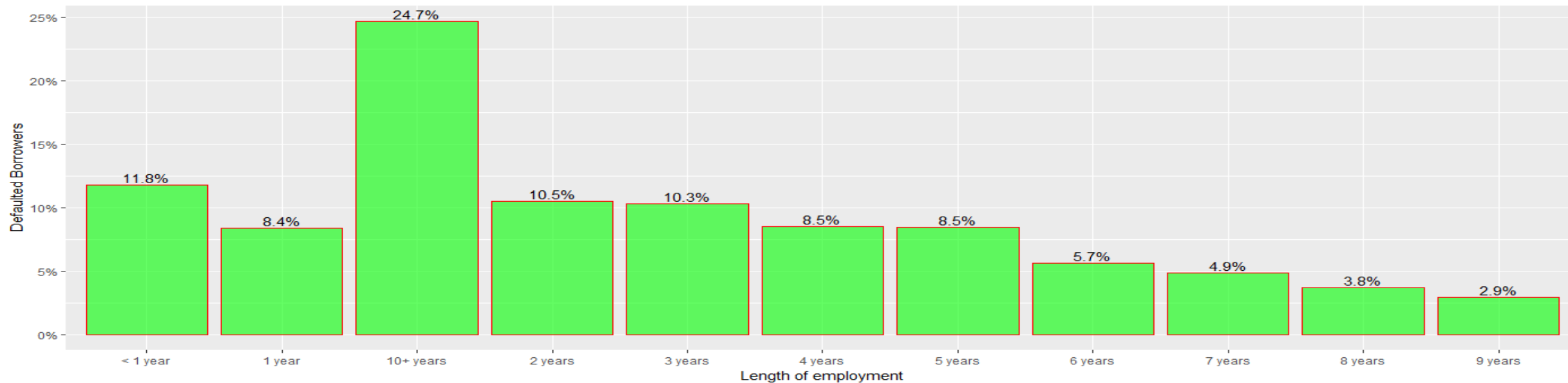
Defaulted loan - grade



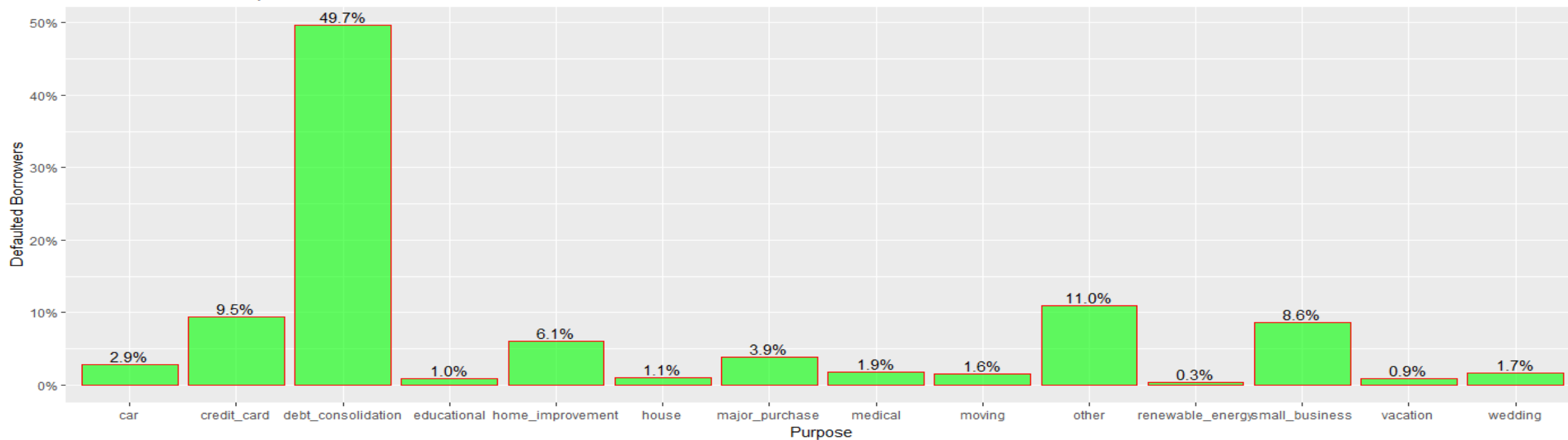
Defaulted loans - Verification Status



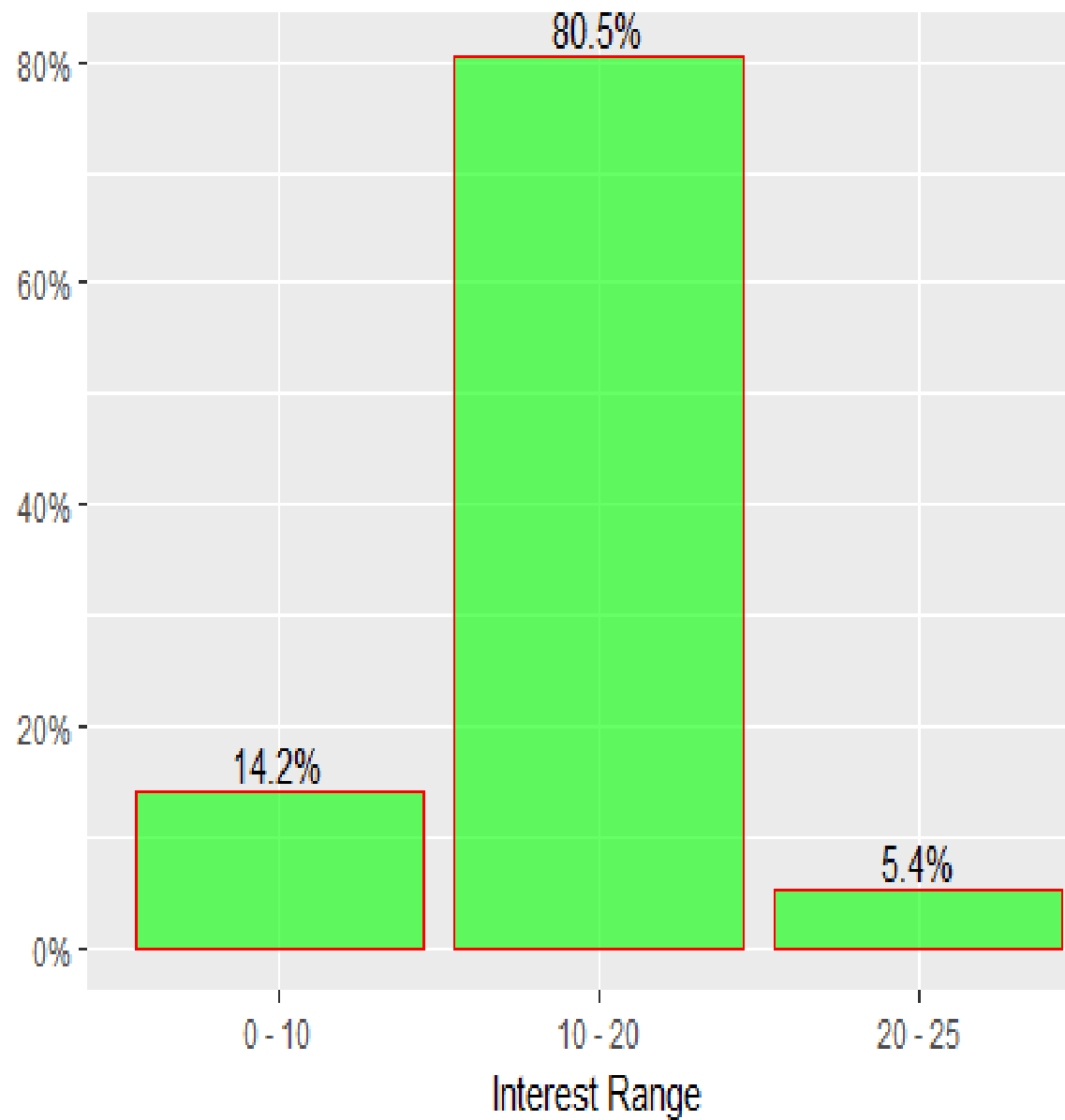
Defaulted loan - Length Of Employment



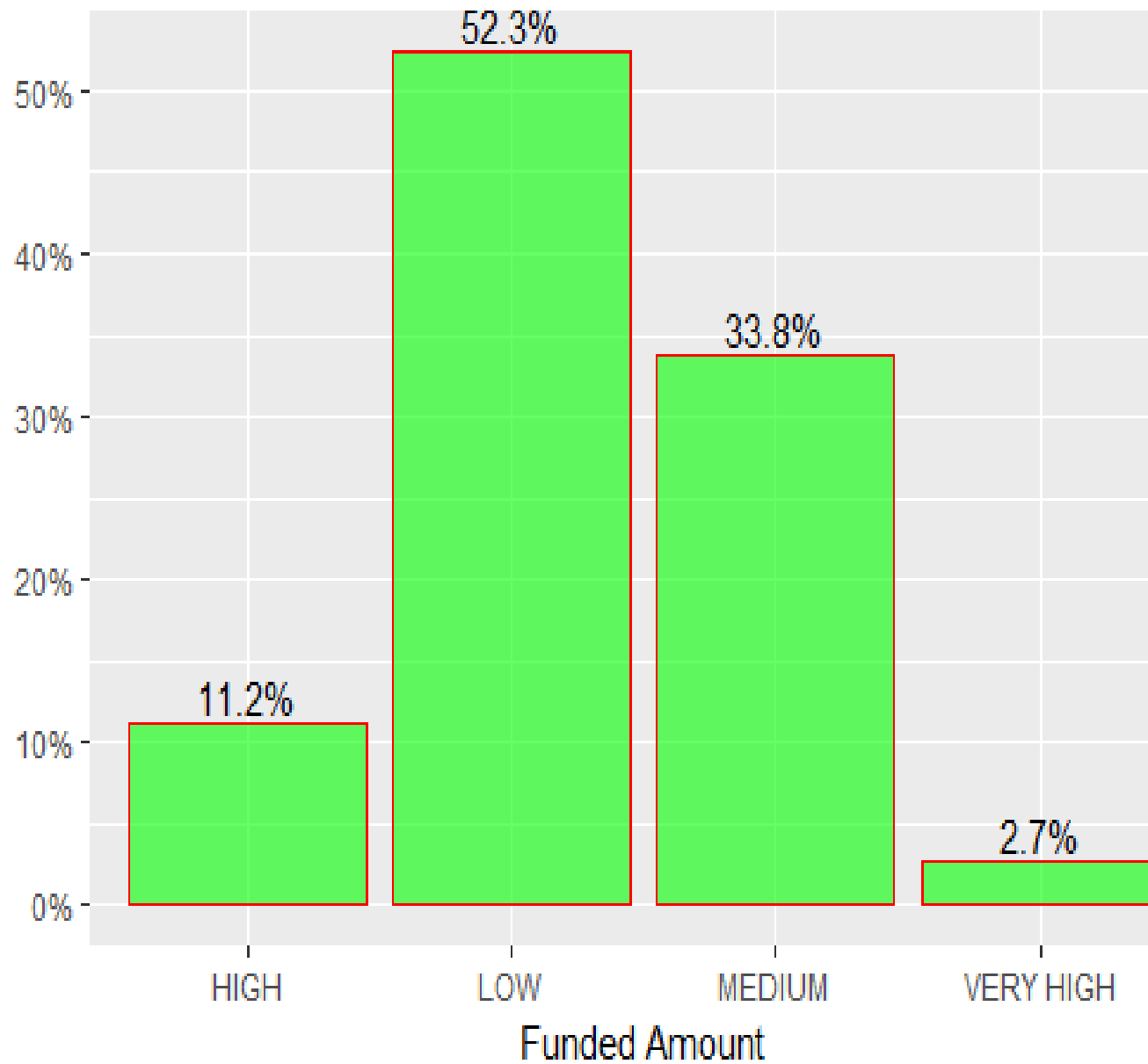
Defaulted loans - Purpose



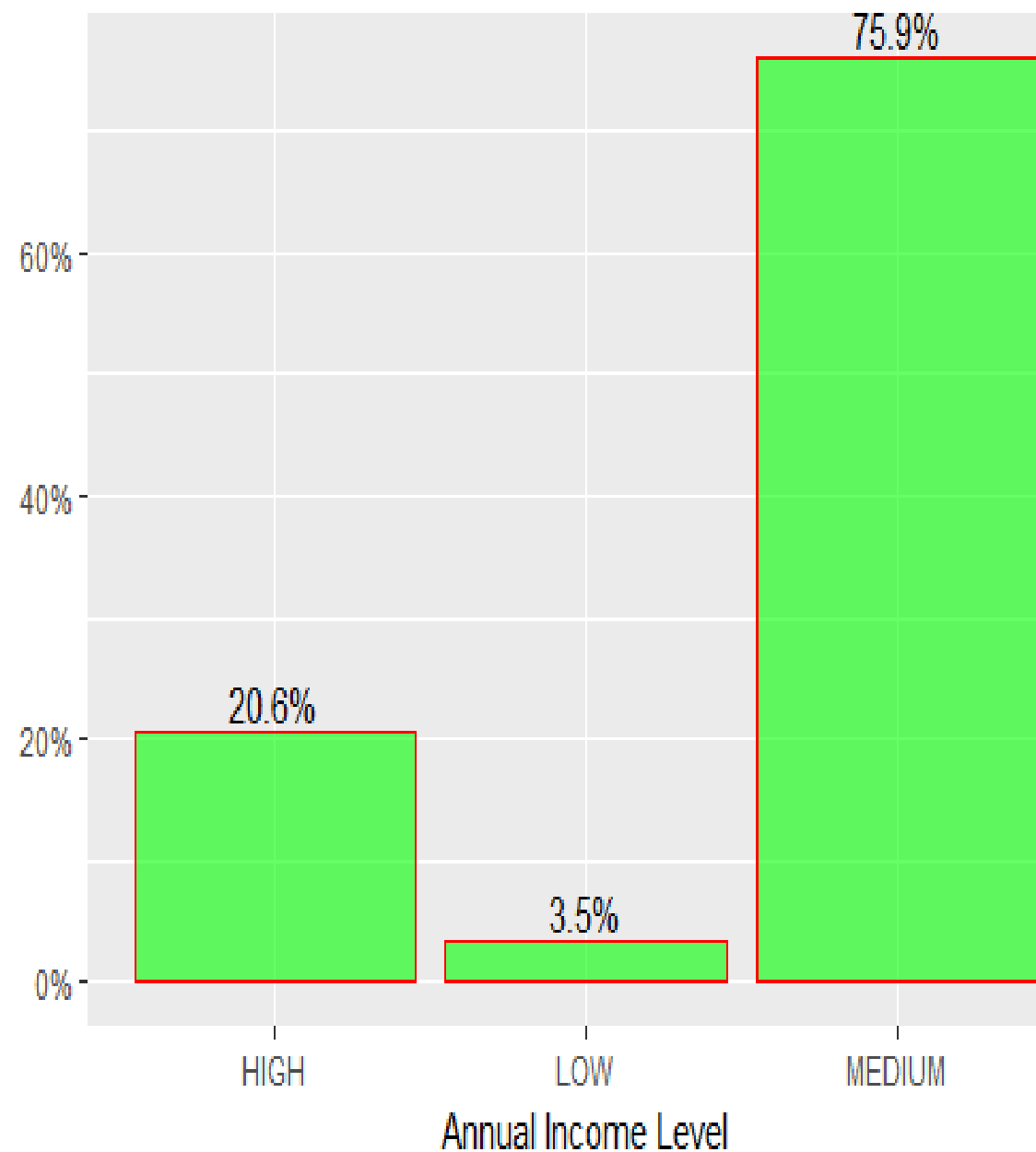
Defaulted loan - Interest Range



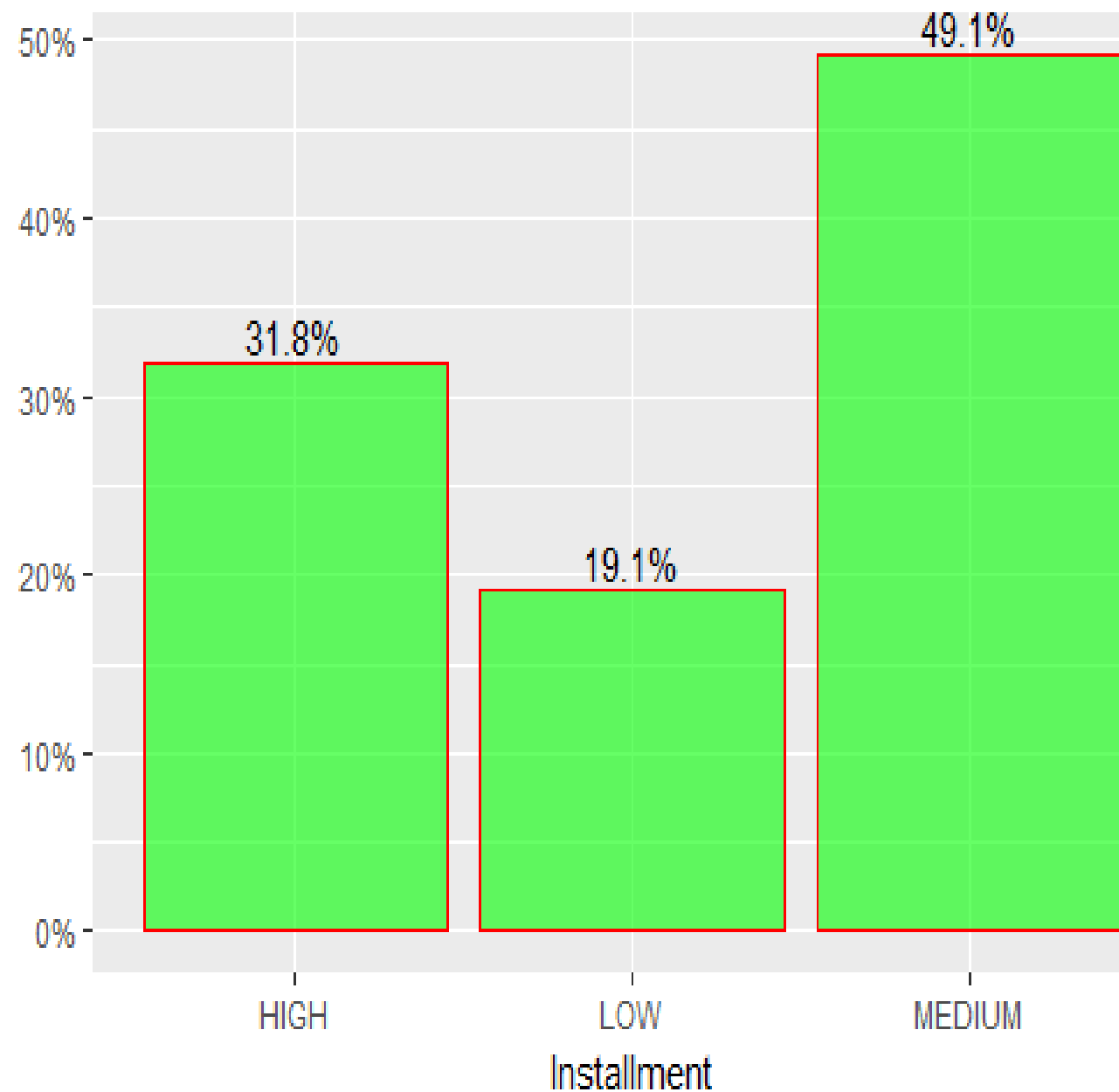
Defaulted loan - Funded Amount



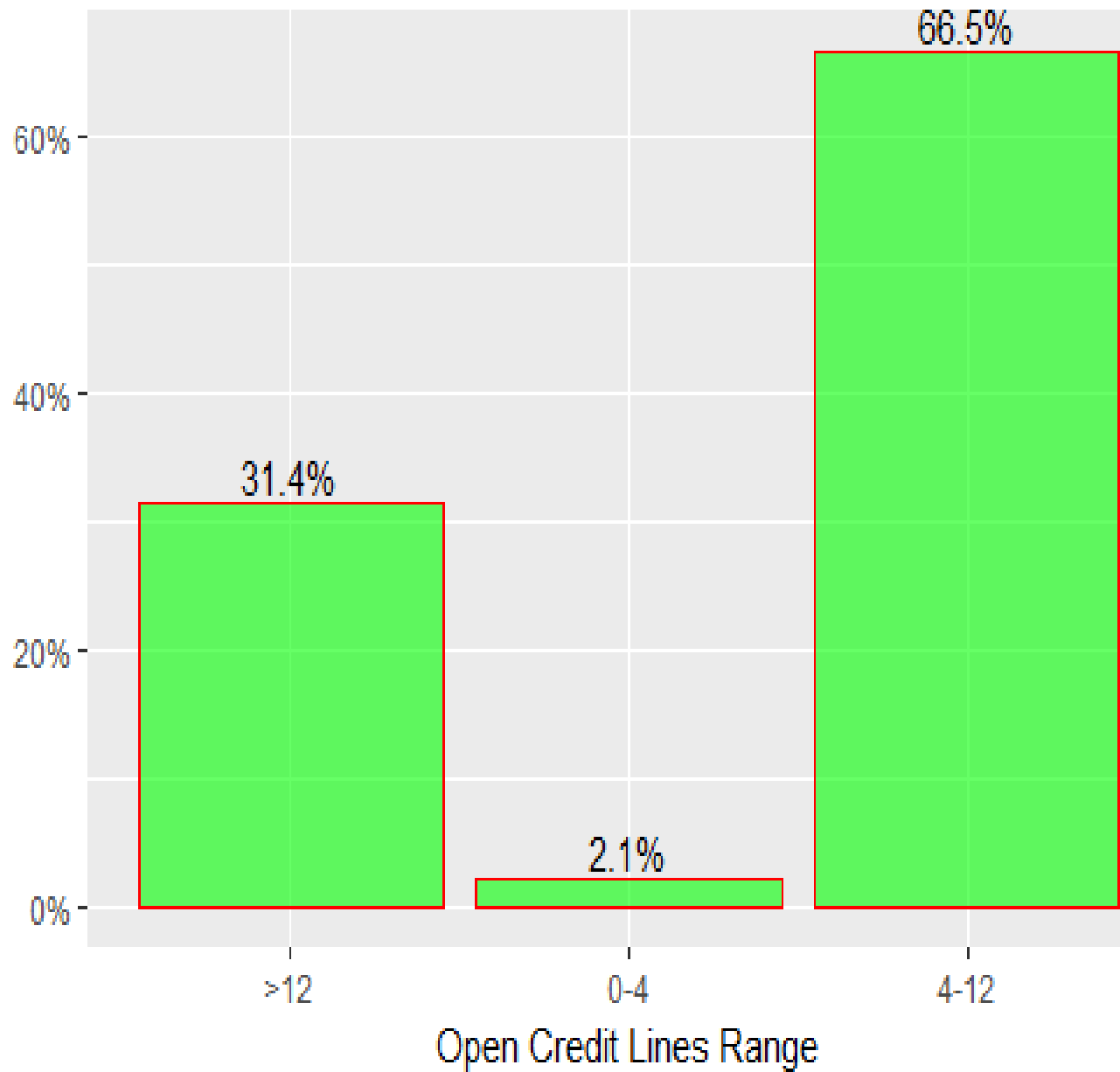
Defaulted loan - Annual Income Level



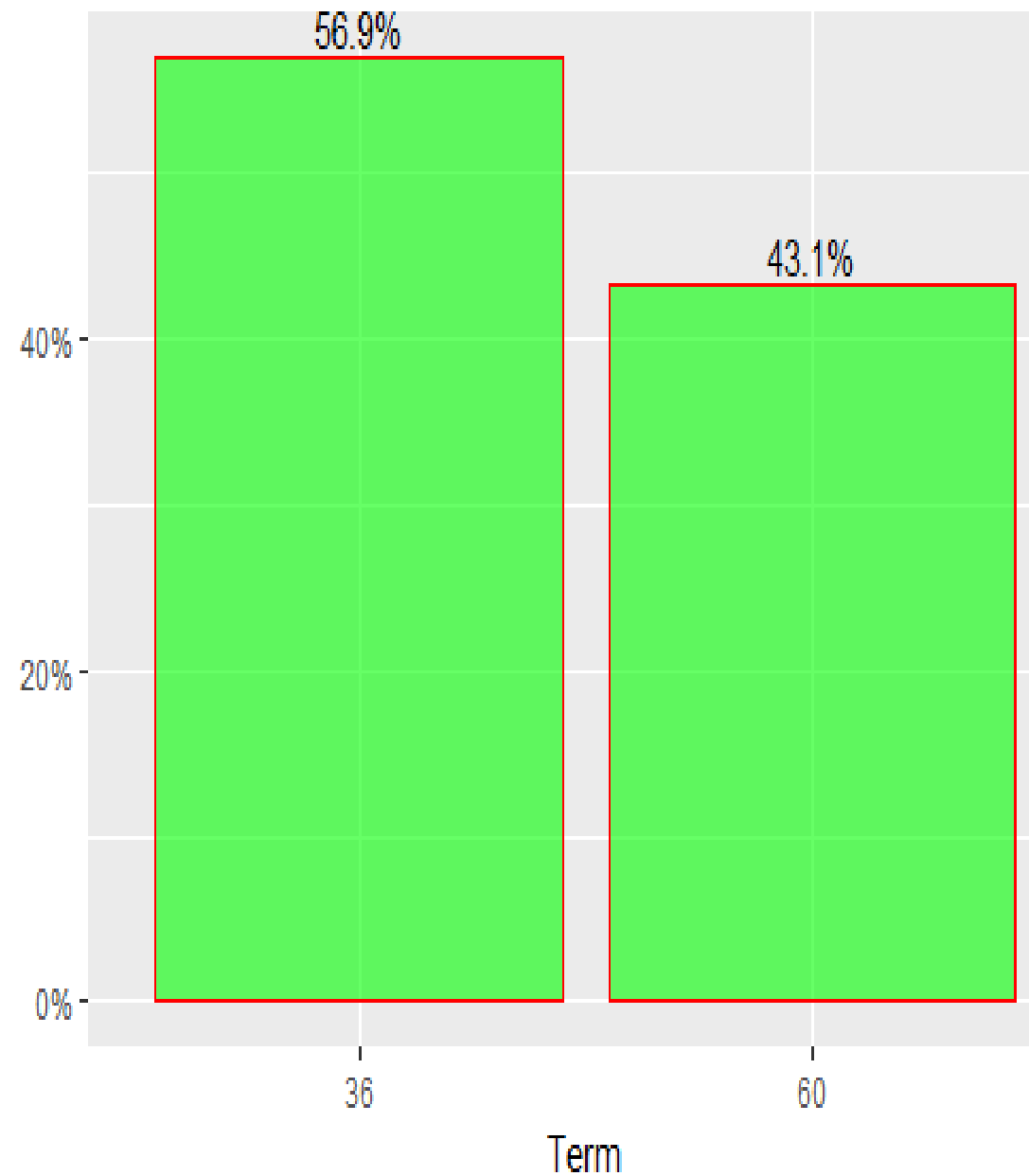
Defaulted loan - Installment Level



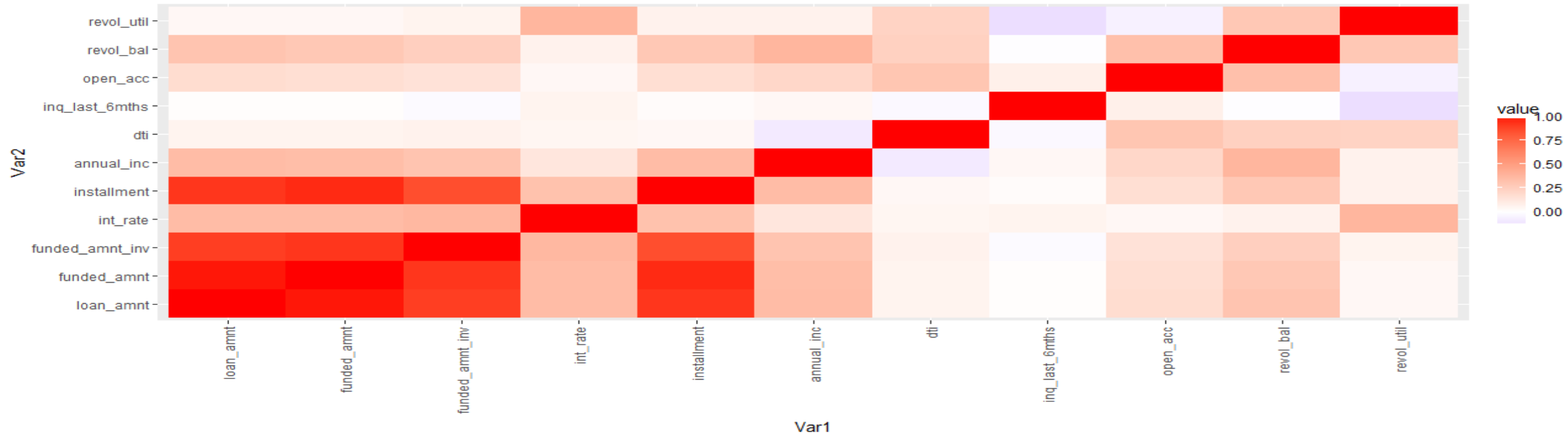
Defaulted loan - Open Credit Lines



Defaulted loan - Term



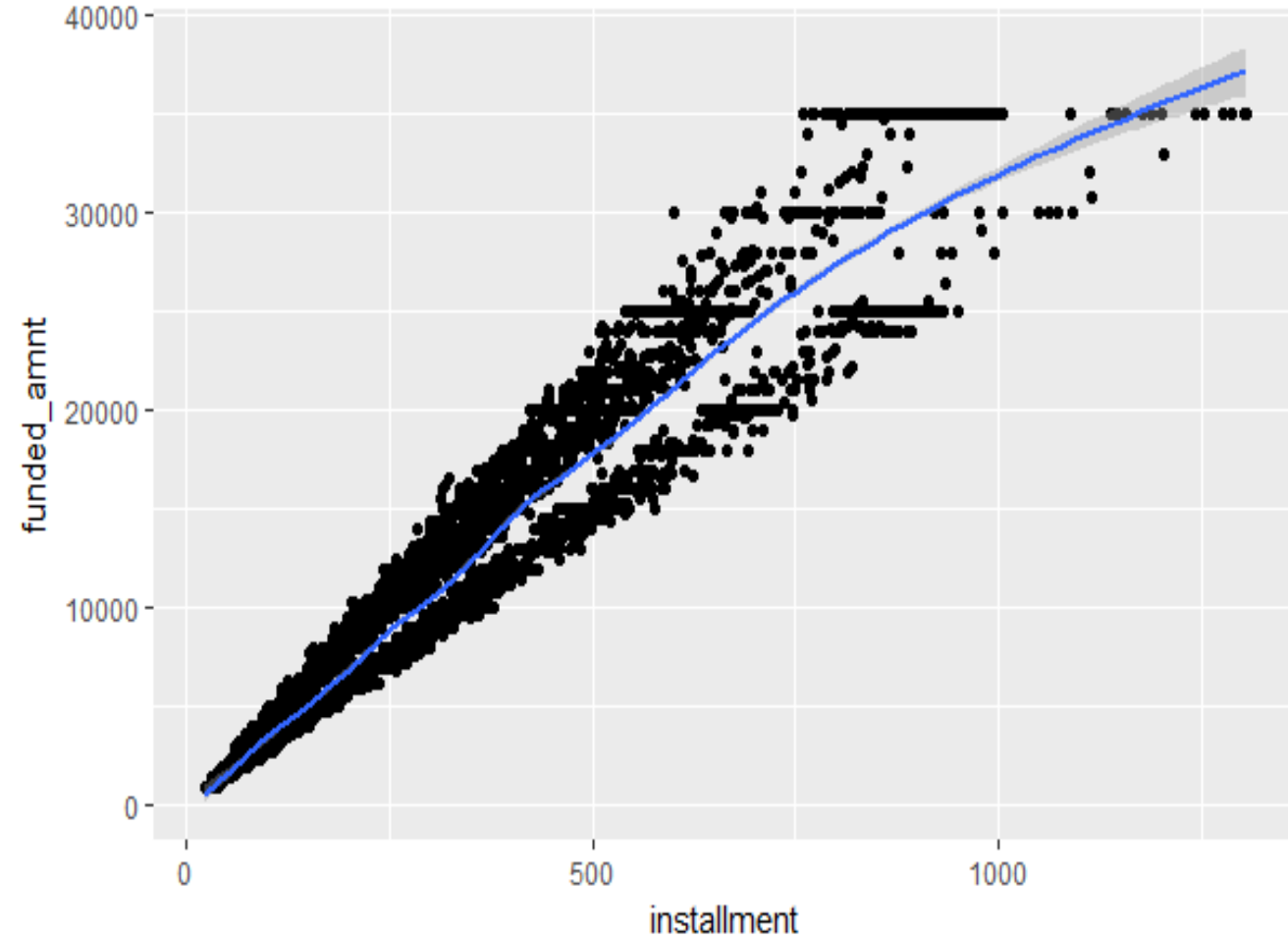
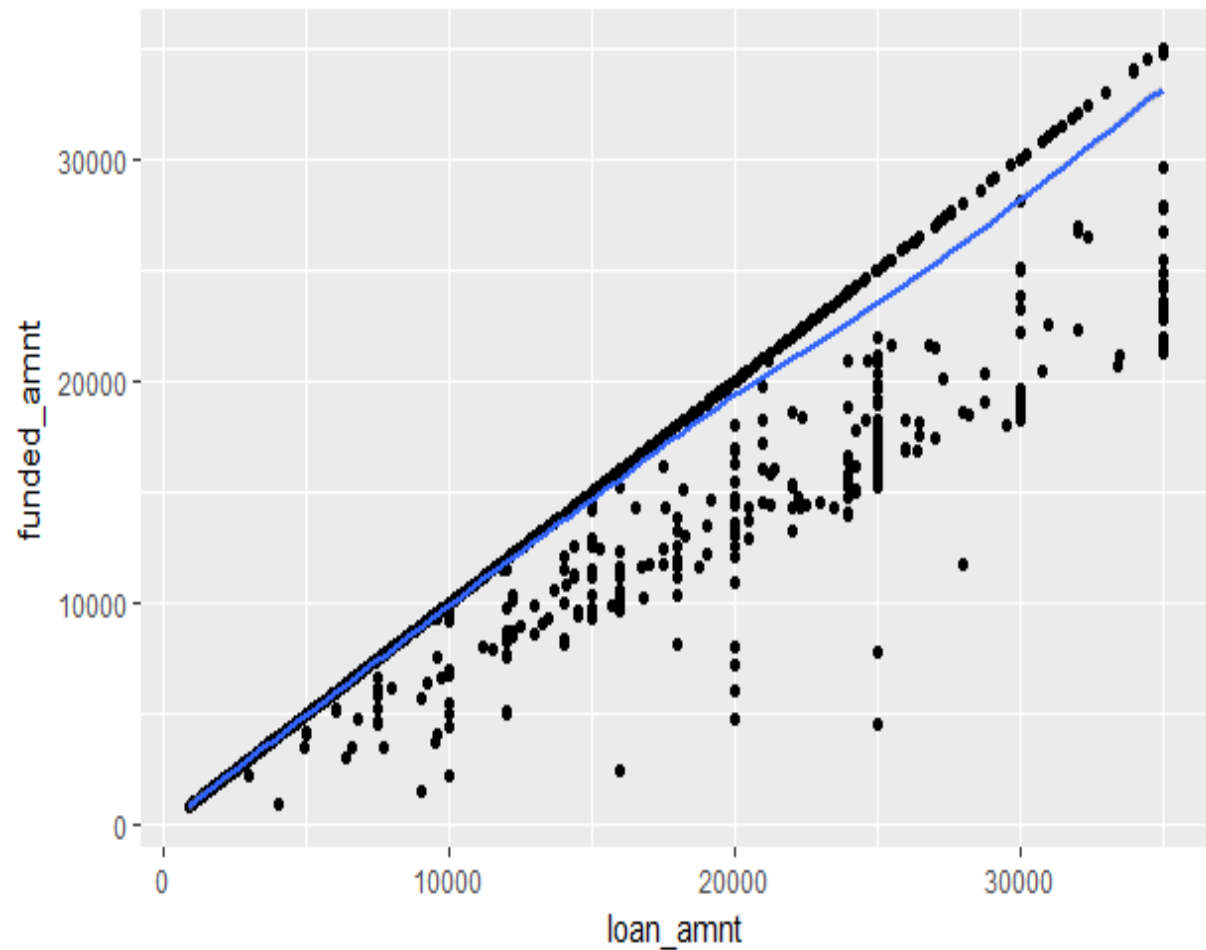
Correlation between Continuous variables



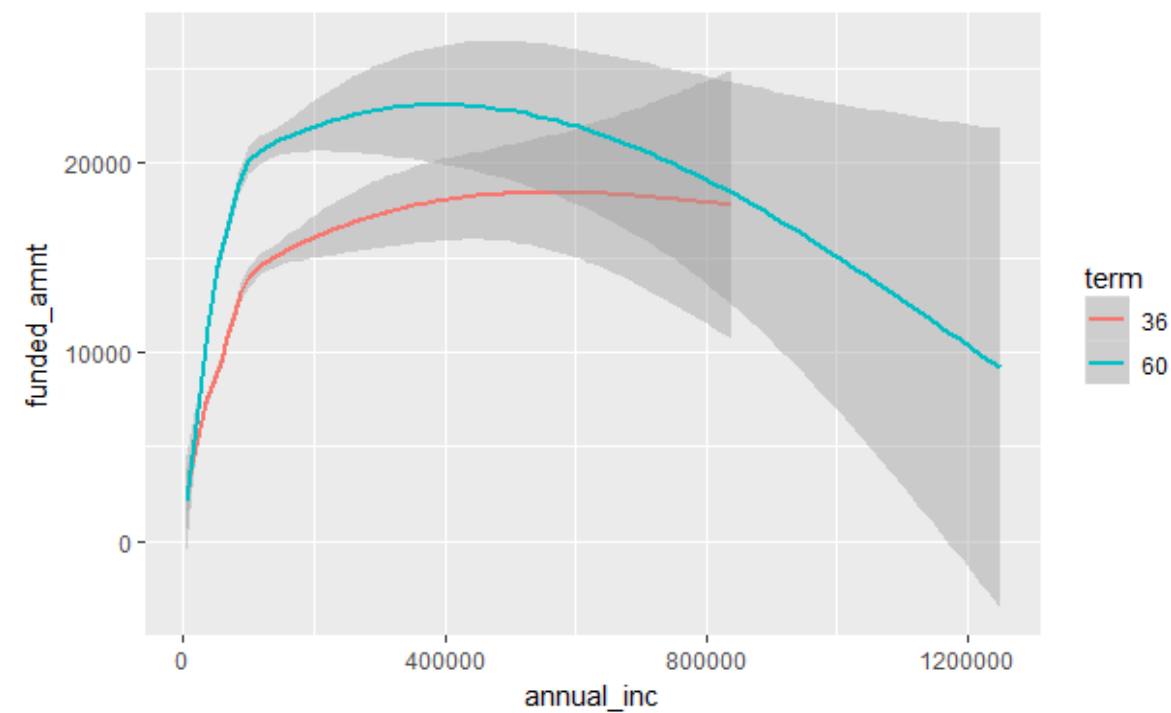
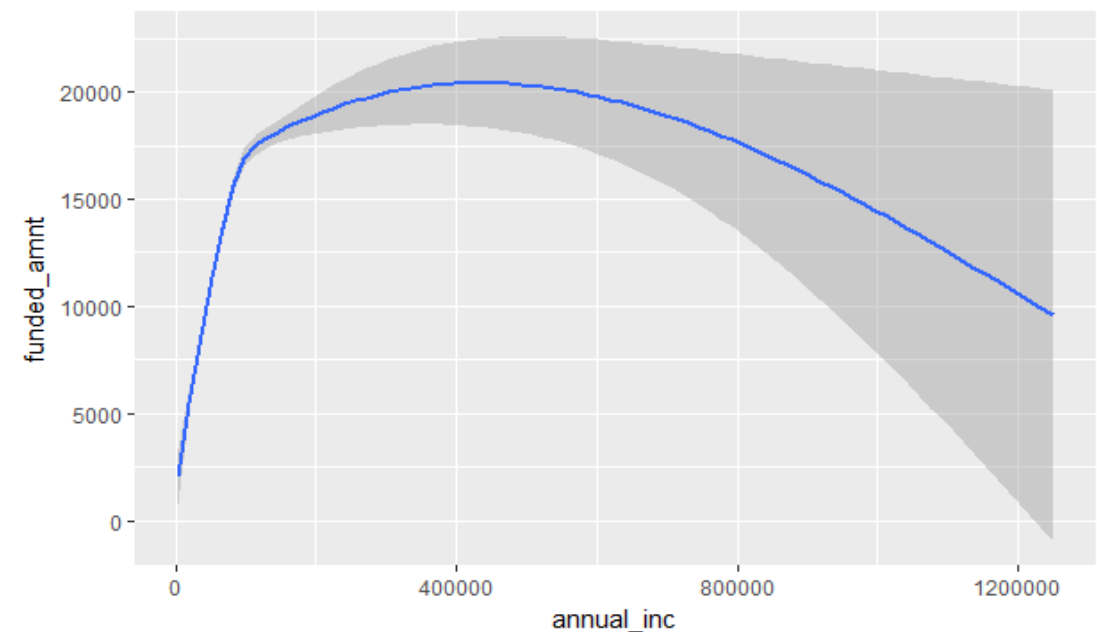
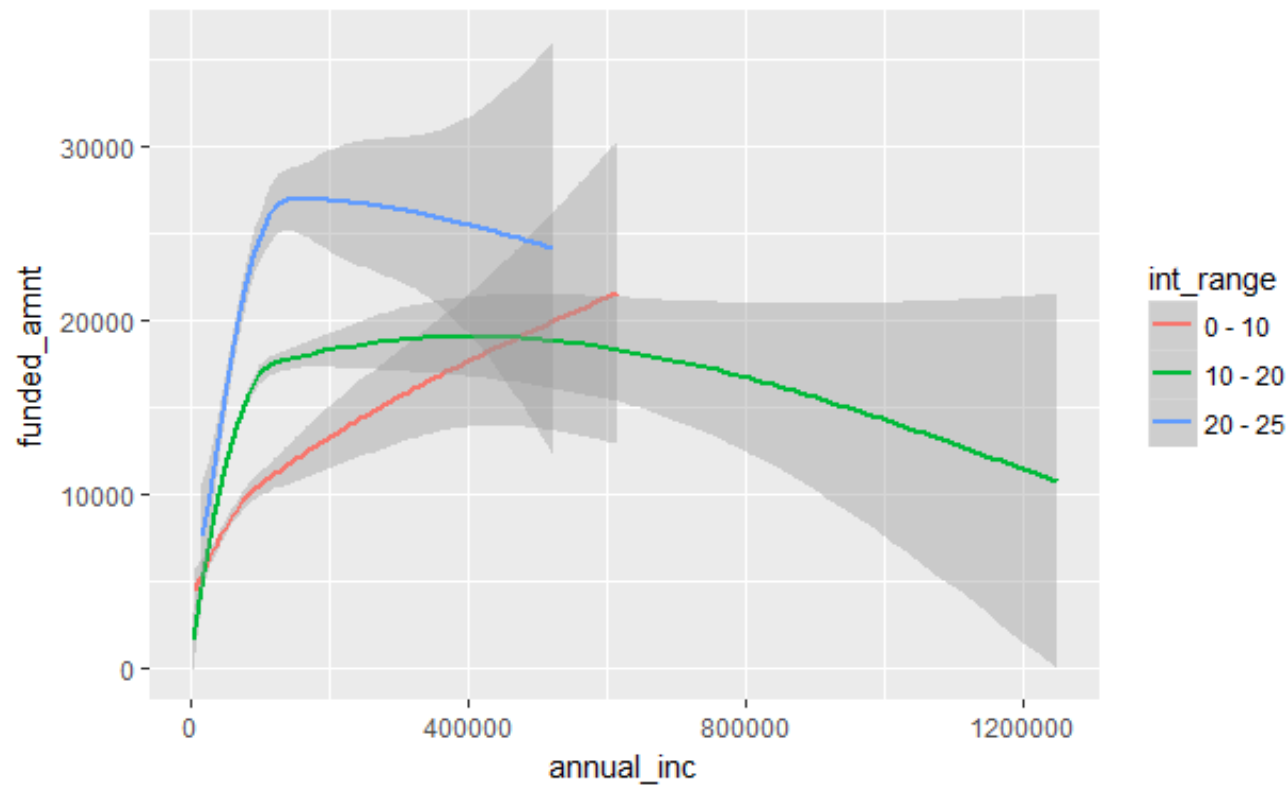
INFERENCES

- Correlation between loan amount and funded amount = 0.97 – Highly correlated
- Correlation between loan and funded amount invested = 0.90 – Highly correlated
- Correlation between loan amount and installment = 0.90 – Highly correlated
- Correlation between loan amount and annual income = 0.34 – **Low correlation**

Funded_amnt vs loan_amnt & Funded_amnt vs installment



Funded_amnt vs (annual_inc) with respect to term and int_range

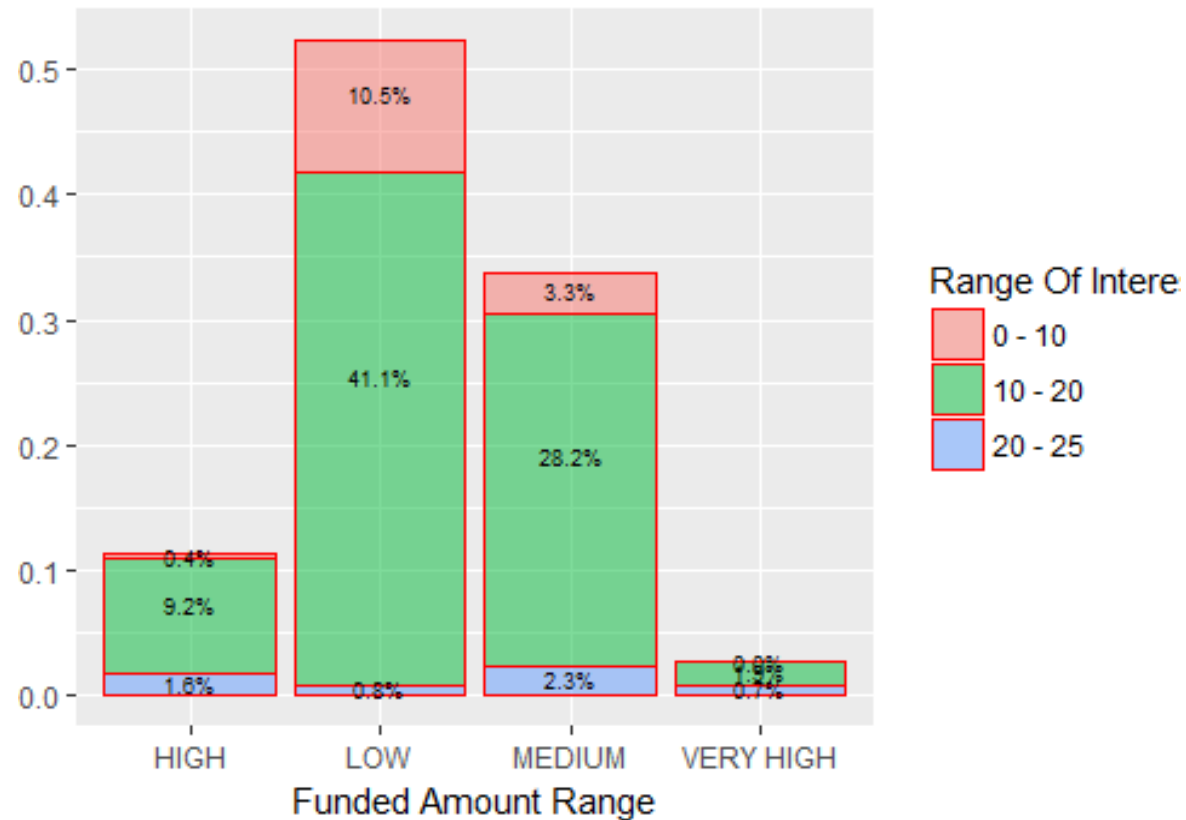


Bivariate Analysis - I

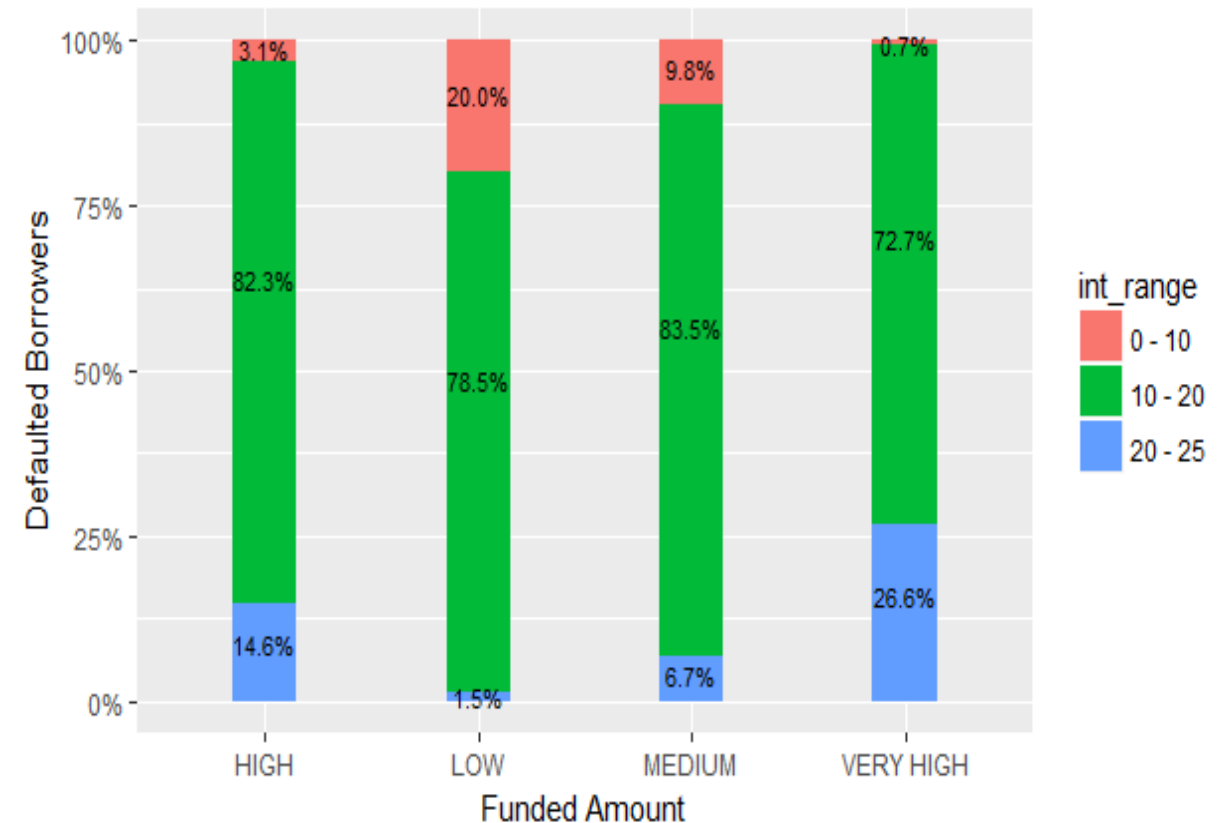
Following are considered for Continuous / Categorical variable

- Funded Amount / Interest Range
- Funded Amount / Term
- Funded Amount / Home Ownership
- Funded Amount / Annual Income Range
- Funded Amount / Installment

Defaulted Loan - Funded Amount / Range of Interest

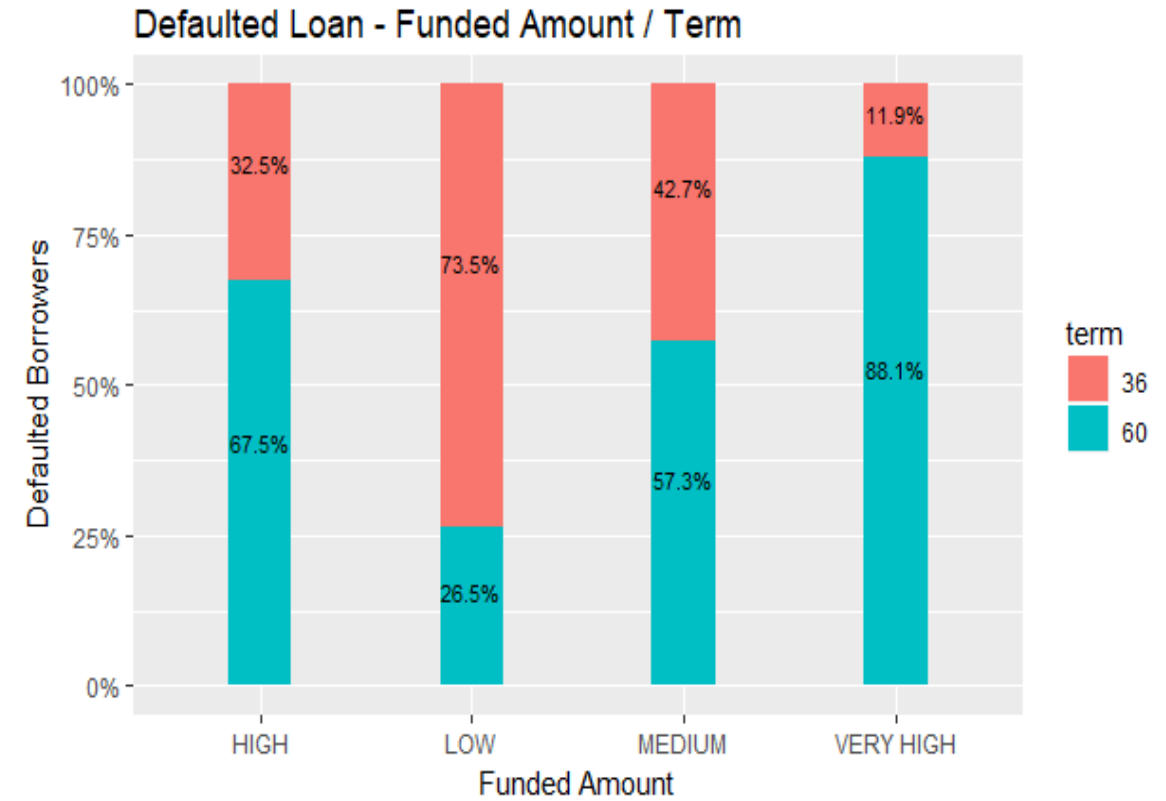
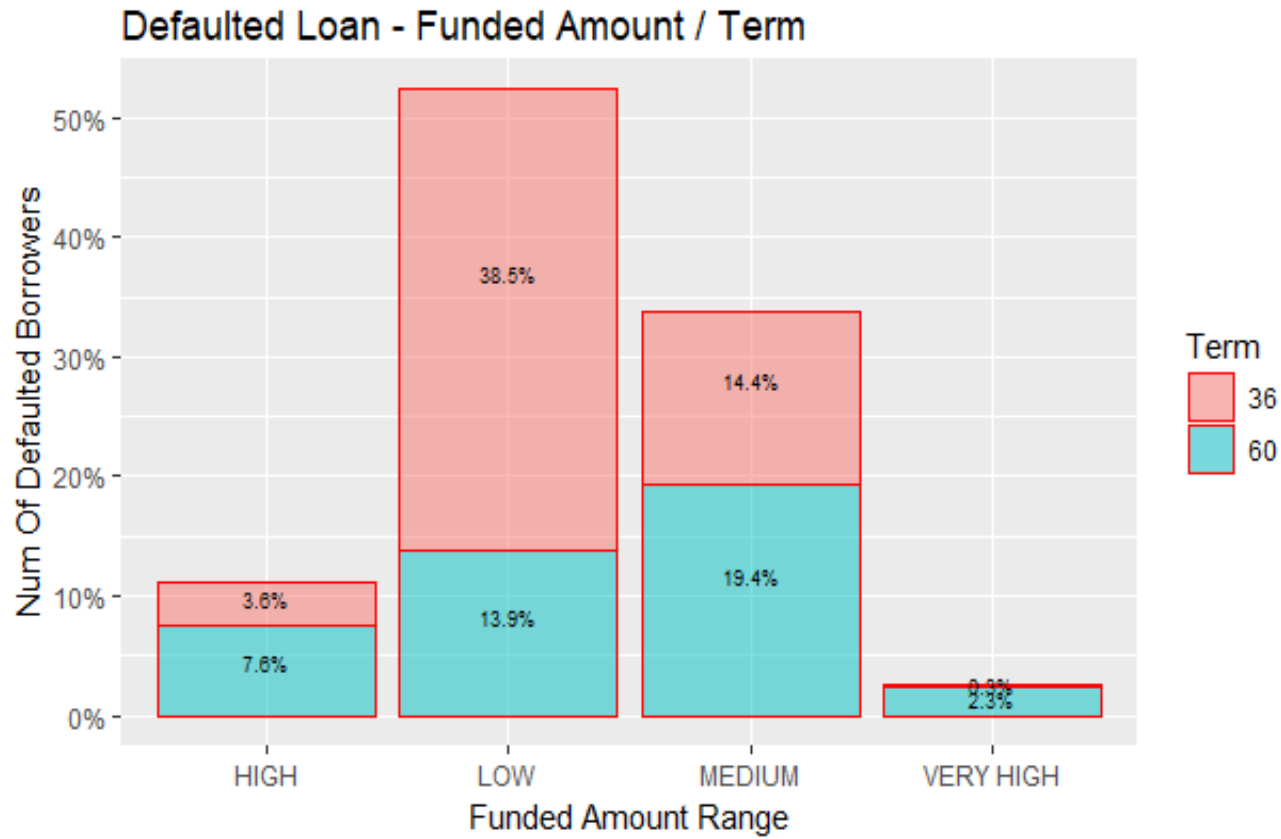


Defaulted Loan - Funded Amount / Range of interest



• INFERENCES:

- Maximum number of loans defaulted in the “LOW RANGE” : 0 – 10k
- Overall : $41.1 + 10.5 + 0.8 = 52.4$ % of loans that are defaulted are in the “LOW” Range i.e. : 0 – 10K
- Out of those loans : 78.5 % are defaulted in the interest range on 10– 20 %



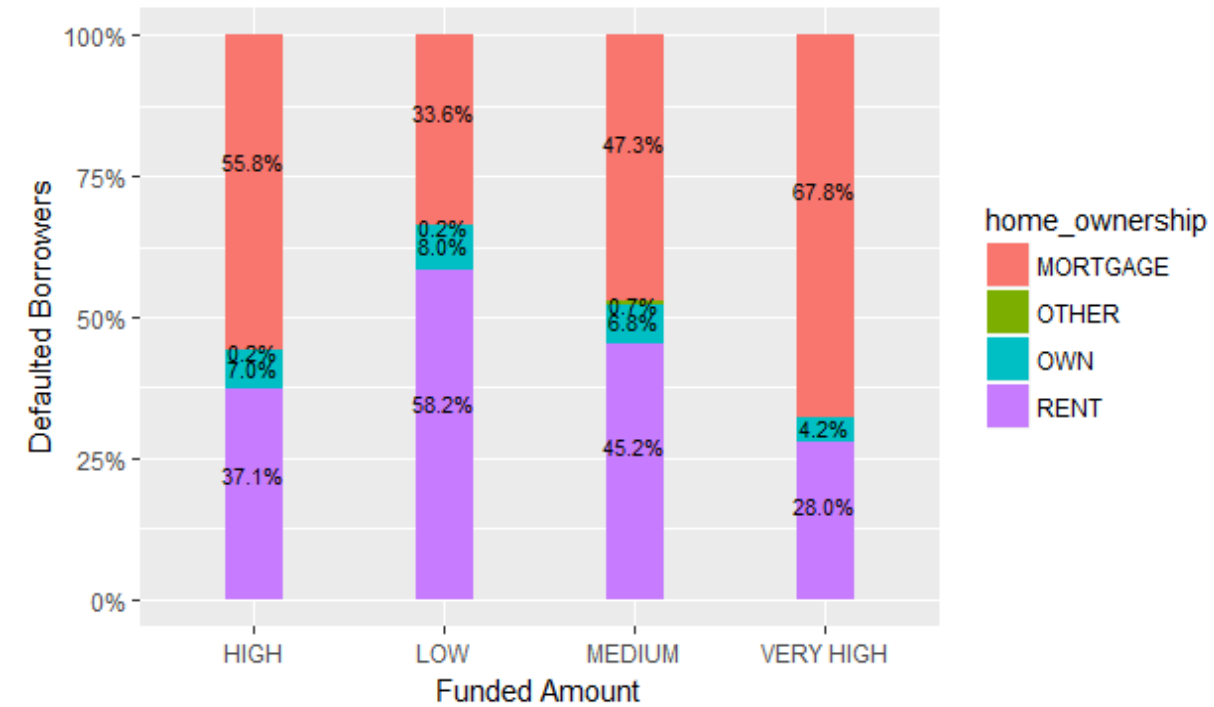
- INFERENCES:**

- Maximum number of loans defaulted in the “LOW RANGE” : 0 – 10k
- Overall : $38.5 + 13.9 = 52.4\%$ of loans that are defaulted are in the “LOW” Range i.e. : 0 – 10K
- Out of those loans : 73.5 % are defaulted that have term of 36 months

Defaulted Loan - Funded Amount / Home Ownership



Defaulted Loan - Funded Amount / Home Ownership



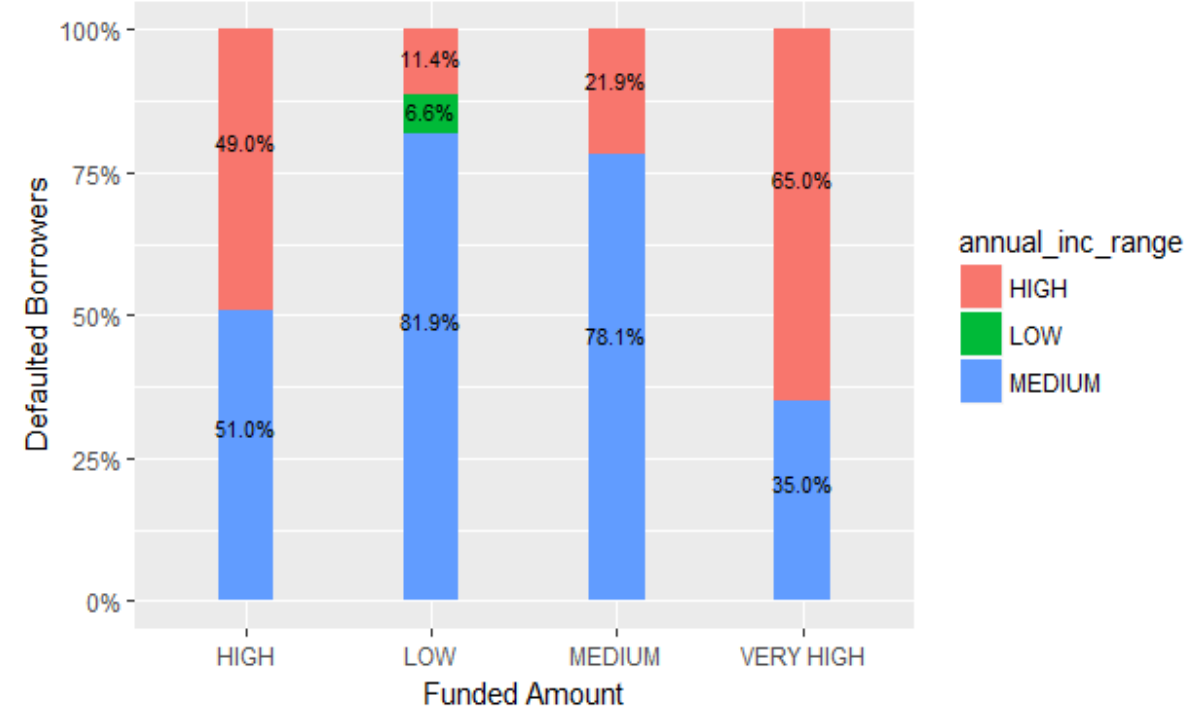
• INFERENCES:

- Maximum number of loans defaulted in the “LOW RANGE” : 0 – 10k
- Overall : 52.4 % of loans that are defaulted are in the “LOW” Range i.e. : 0 – 10K
- Out of those loans :
 - 58.2 % loans are defaulted by those borrowers who are having home status : “RENT”
 - 33.6 % loans are defaulted by those borrowers who are having home status : “MORTGAGE”

Defaulted Loan - Funded Amount Range / Annual Income Level

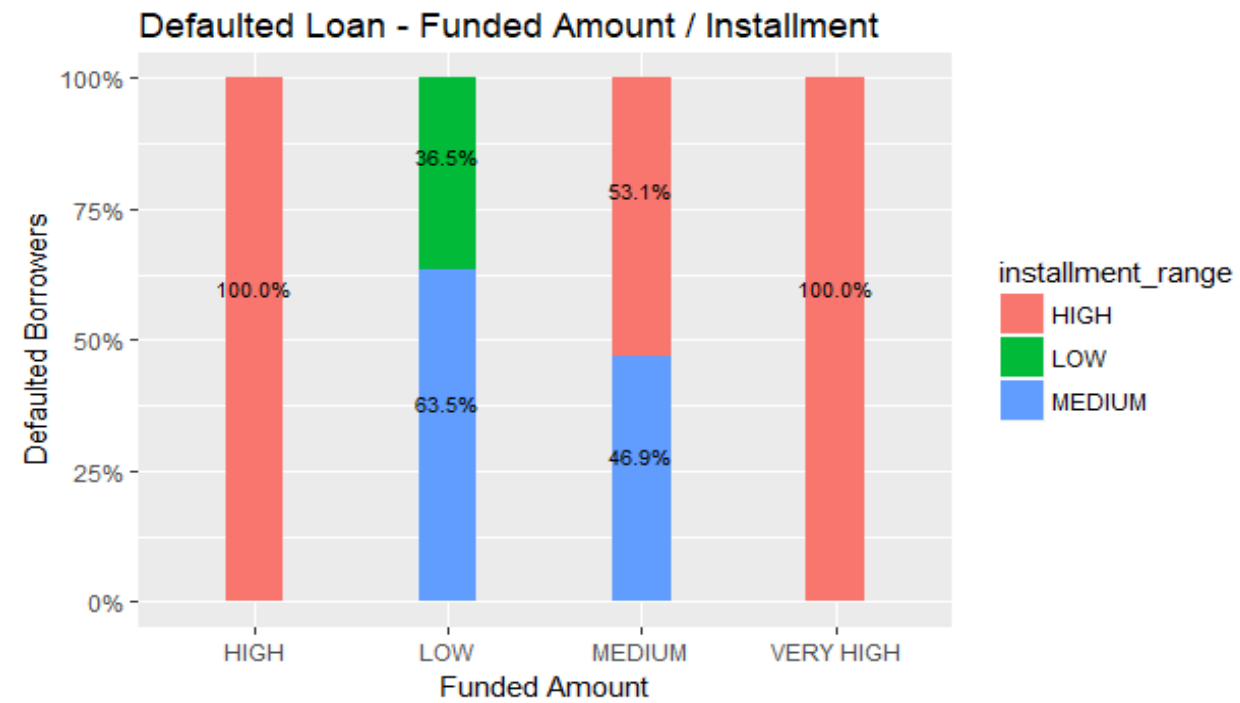
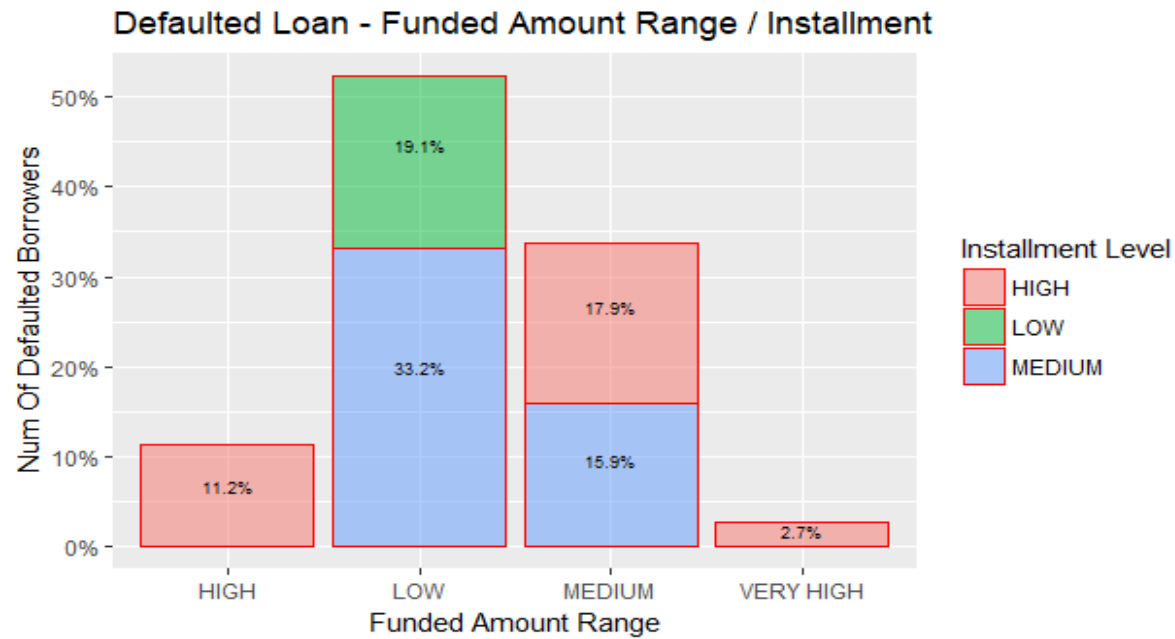


Defaulted Loan - Funded Amount / Annual Income



• INFERENCES:

- Maximum number of loans defaulted in the “LOW RANGE” : 0 – 10k
- Overall : 52.4 % of loans that are defaulted are in the “LOW” Range i.e. : 0 – 10K
 - 42.9% - MEDIUM INCOME LEVEL : 20K – 80K
 - 3.5 % - LOW INCOME LEVEL : 0 -20K
 - 6 % - HIGH INCOME LEVEL : > 80K
- Out of those loans :
 - 81.9 % loans are defaulted by those borrowers who are in MEDIUM INCOME LEVEL : 20K – 80K



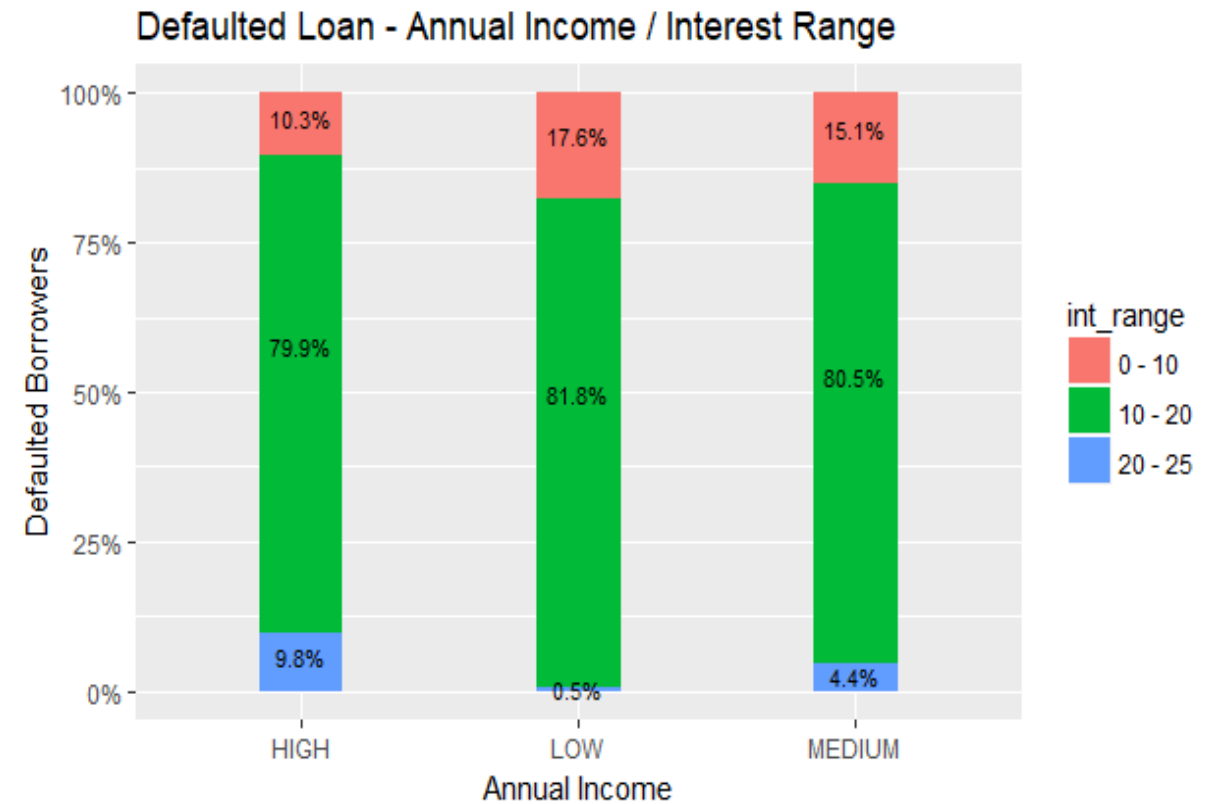
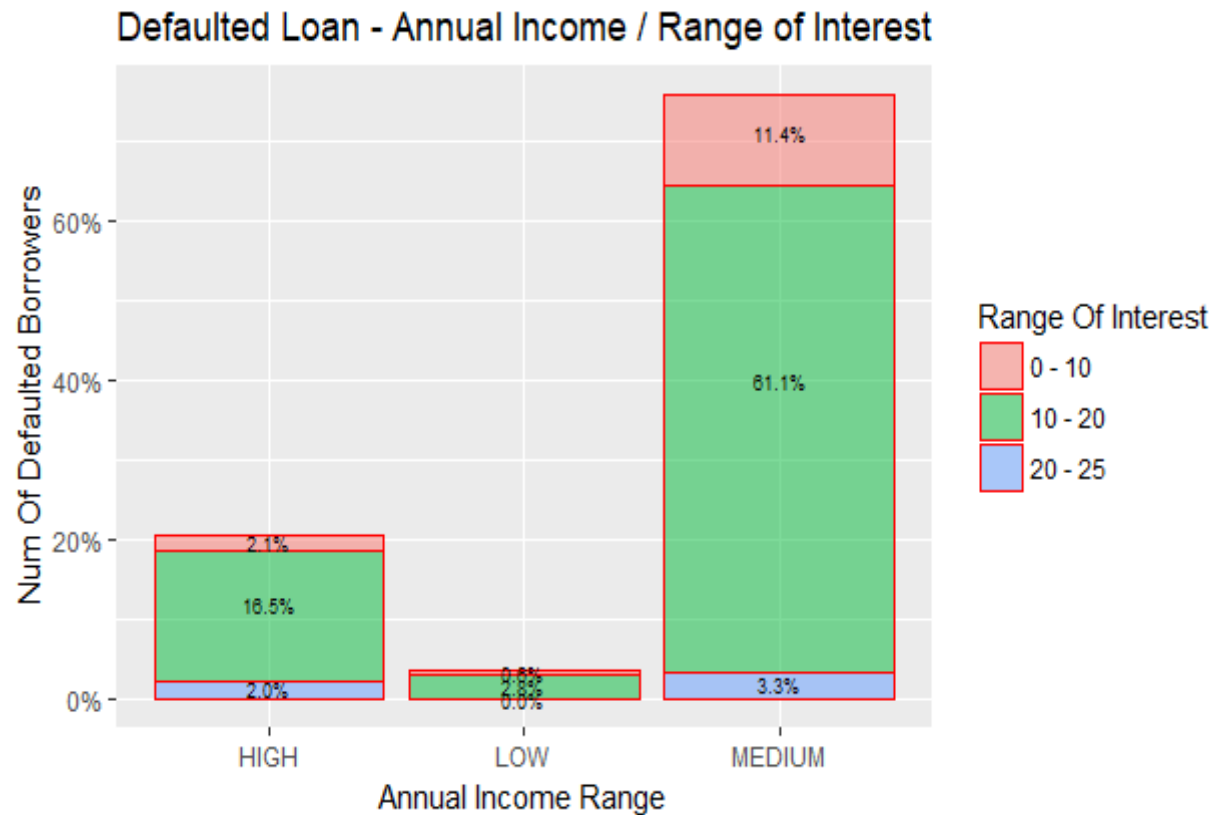
- INFERENCES:**

- Maximum number of loans defaulted in the “LOW RANGE” : 0 – 10k
- Overall : 52.4 % of loans that are defaulted are in the “LOW” Range i.e. : 0 – 10K
 - 33.2 % MEDIUM INSTALLMENT RANGE : 150 – 400
 - 19.1 % LOW INSTALLMENT RANGE : 0 - 150
- Out of those loans :
 - 63.5 % loans are defaulted by those borrowers who are in MEDIUM INCOME LEVEL : 150 - 400

Bivariate Analysis - II

Following are considered for Continuous / Categorical variable

- Annual Income Range / Interest Range
- Annual Income Range / Term
- Annual Income Range / Home Ownership
- Annual Income Range / Installment



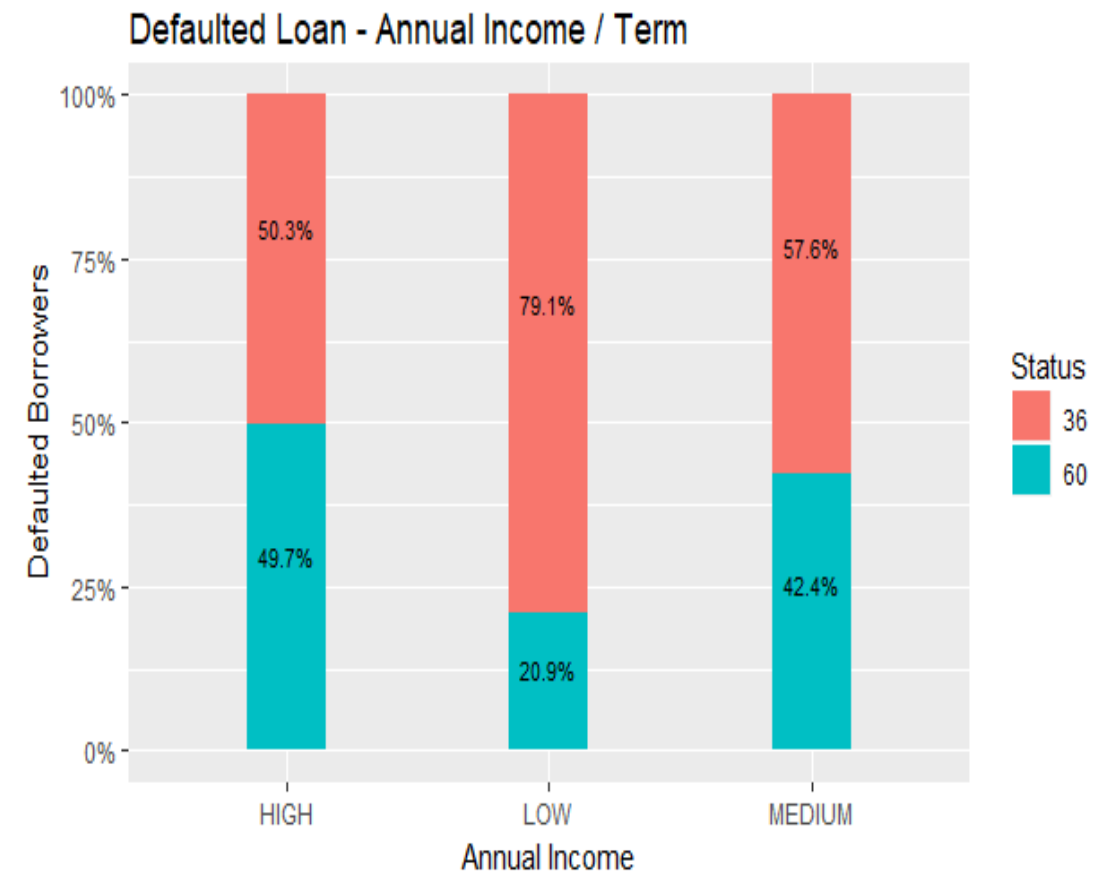
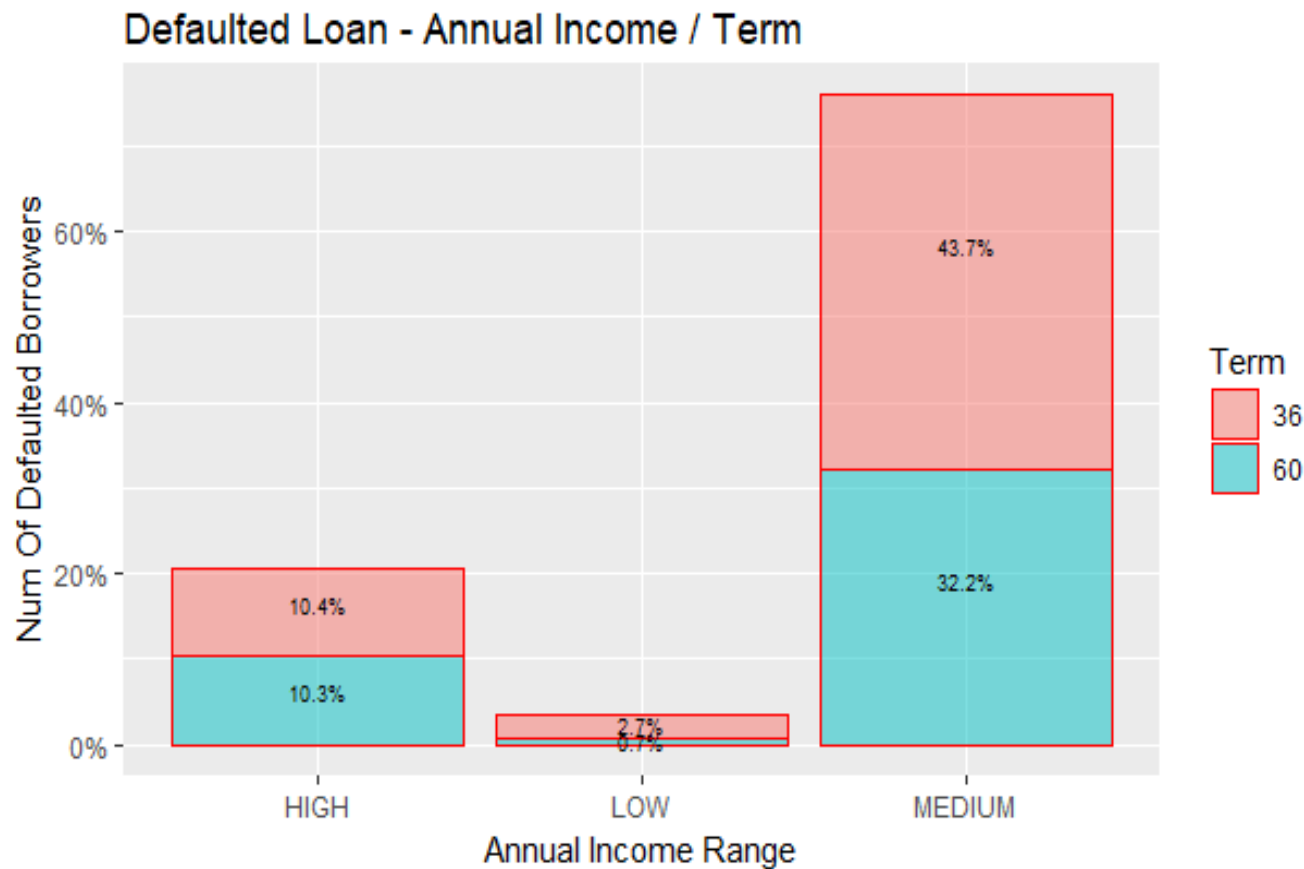
INFERENCES:

Maximum no of loans defaulted by borrowers who are in Medium income level.

Overall : 75.8 % of loans that are defaulted are in the “MEDIUM” Income level i.e. : 20K – 80K

Out of those loans :

80.5 % loans are defaulted in the interest range of 10 – 20 %



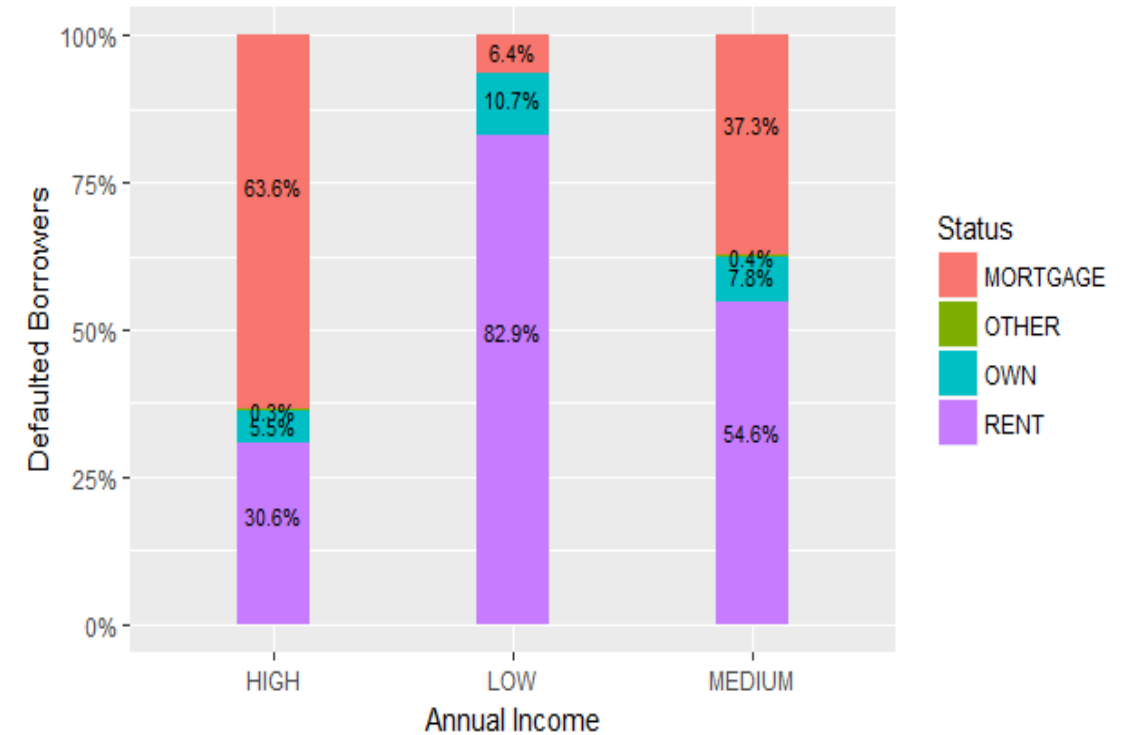
- **INFERENCES:**

- Maximum no of loans defaulted by borrowers who are in Medium income level
- Overall : 75.8 % of loans that are defaulted are in the “MEDIUM” income level i.e. : 20K – 80K
- Out of those loans : 57.8 % are defaulted that have term of 36 months

Defaulted Loan - Annual Income / Home Ownership



Defaulted Loan - Annual Income / Home Ownership



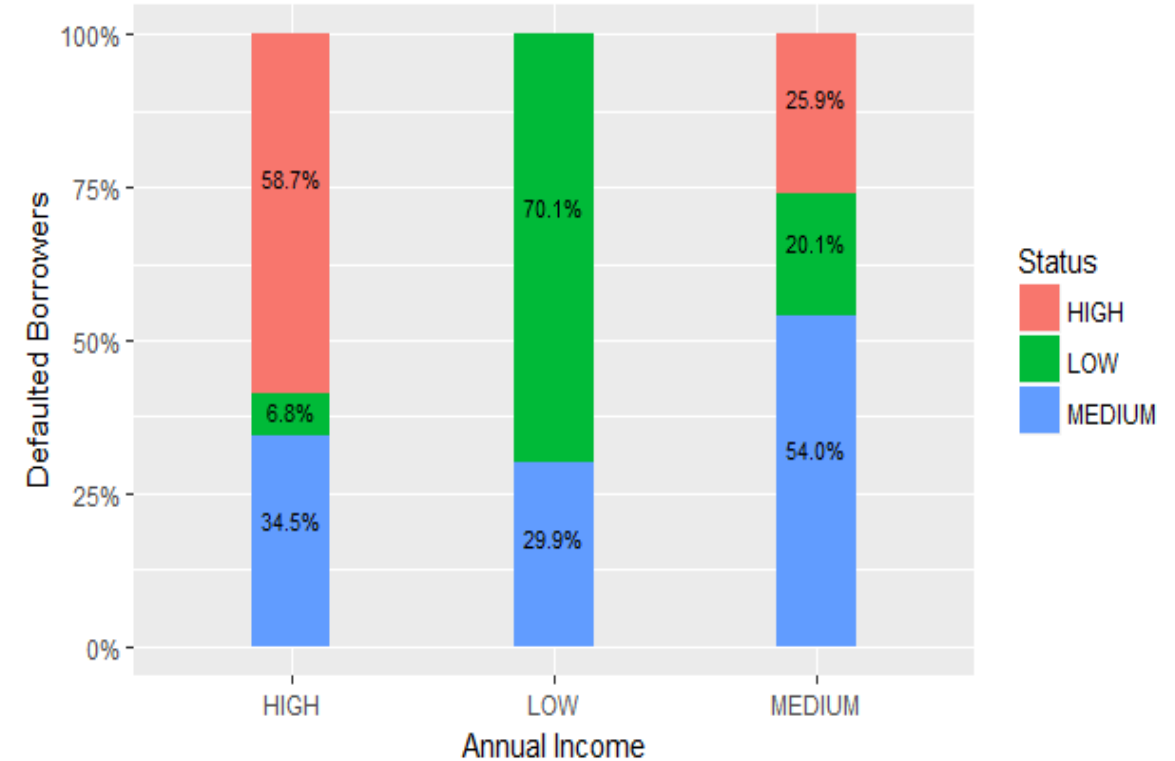
• INFERENCES:

- Maximum no of loans defaulted by borrowers who are in Medium income level
- Overall : 75.8 % of loans that are defaulted are in the “MEDIUM” income level i.e. : 20K – 80K
- Out of those loans :
 - 54.6% of loans defaulted by borrowers who have home_ownership status : RENT
 - 37.3 % of loans defaulted by borrowers who have home_ownership status : MORTGAGE

Defaulted Loan - Annual Income Range / Installment Range



Defaulted Loan - Annual Income / Installment Range



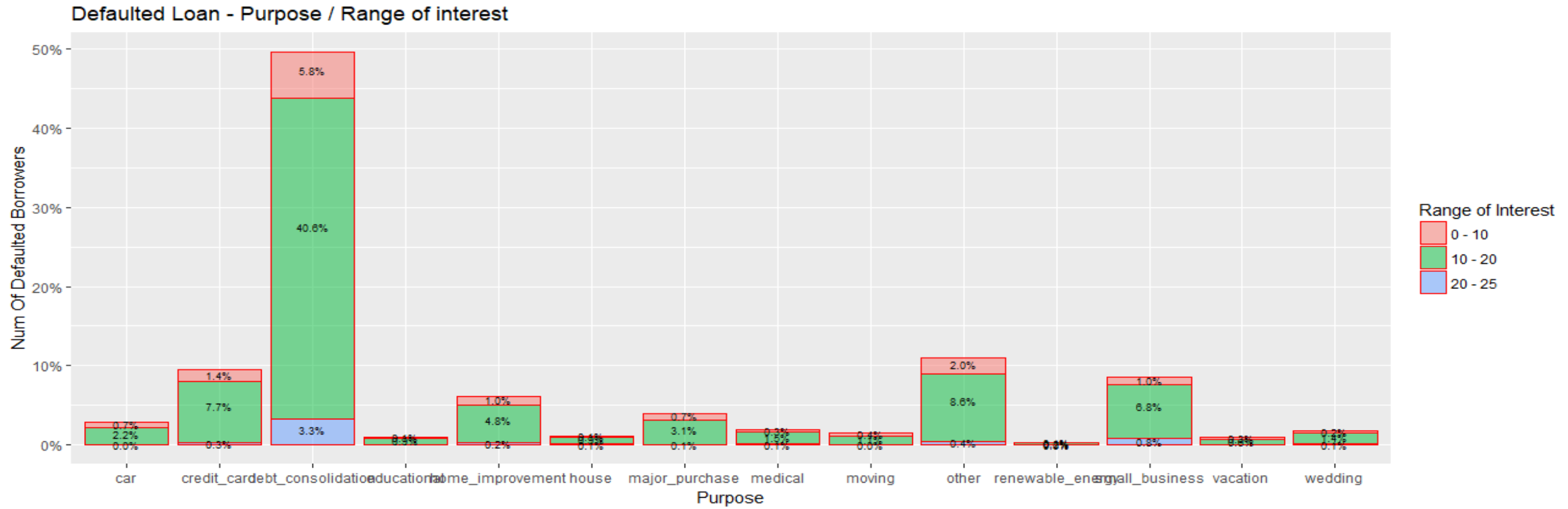
• INFERENCES:

- Overall : 75.8 % of loans that are defaulted are in the “MEDIUM” income level i.e. : 20k – 80k
- Out of those loans :
 - 54 % of loans defaulted by borrowers who have to pay MEDIUM INSTALLMENT : 150 - 400

Bivariate Analysis - III

Following are considered for Continuous / Categorical variable

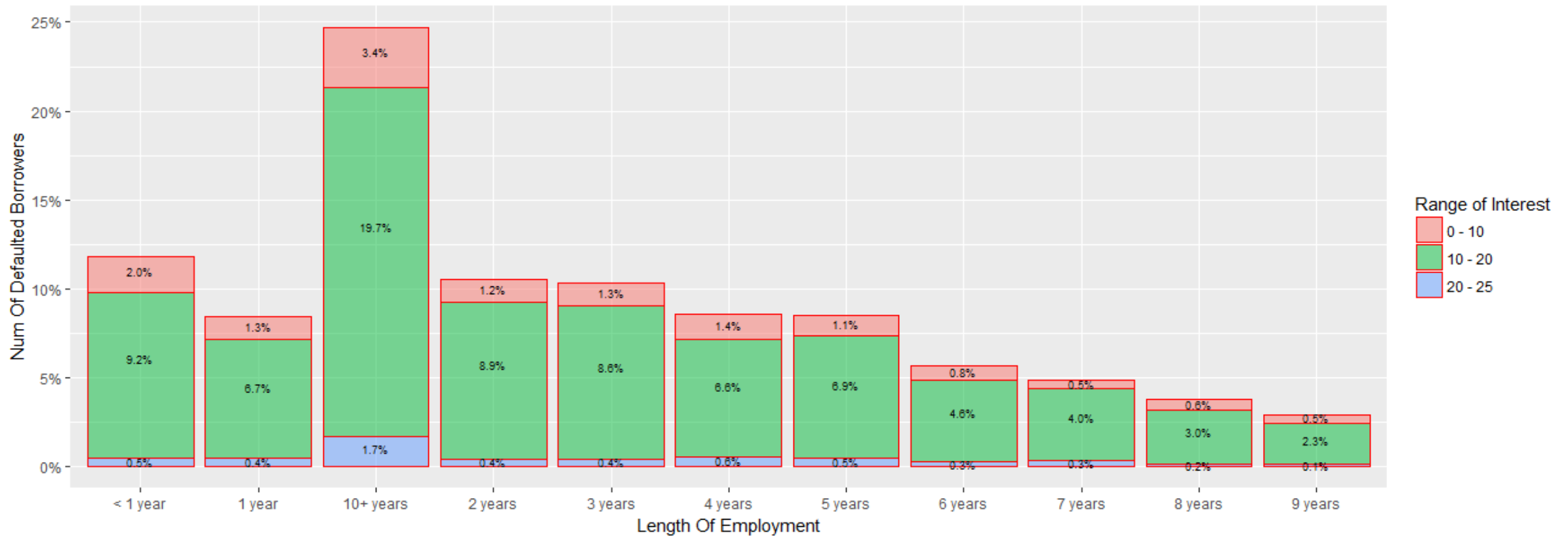
- Purpose / Interest Range
- Employment Length / Interest Range
- Grade / Interest Range
- Employment Length / Purpose
- Open Credit Lines / Term



INFERENCES:

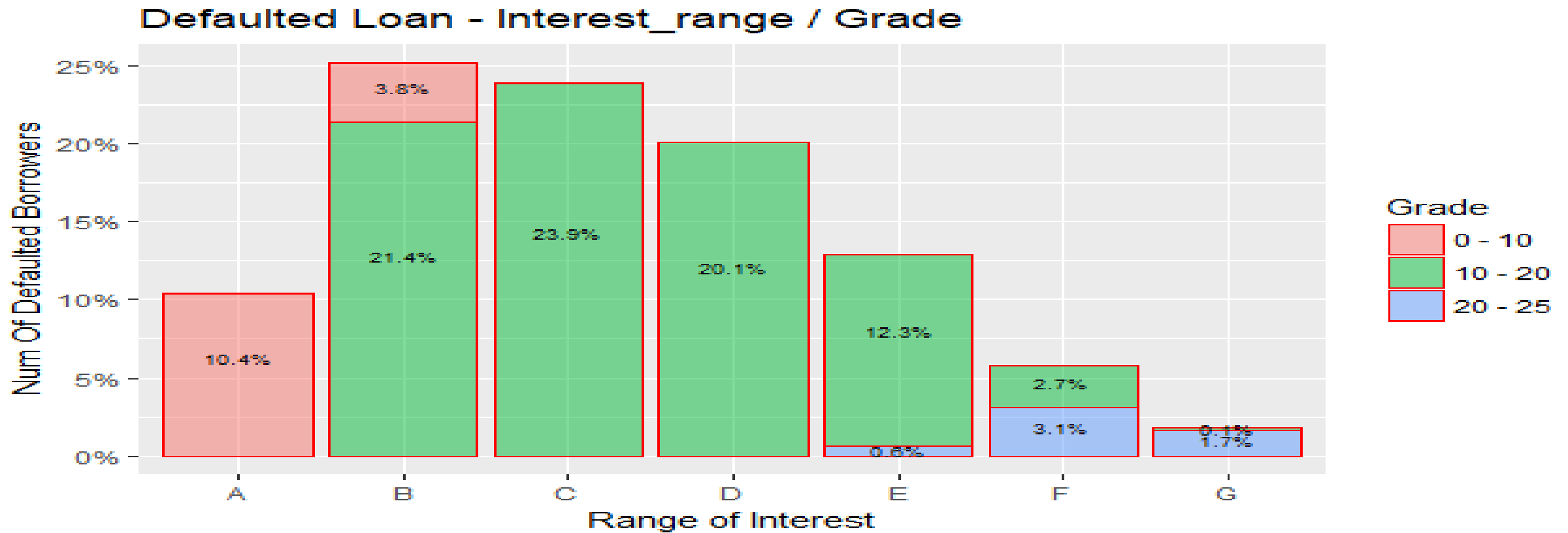
- Overall : 49.7 % of borrowers have taken the loan for debt consolidation:
 - Out of them : 81.7 % of people have taken the loan in the interest range of 10 – 20%

Defaulted Loan - Lenth of employment / Range of interest



INFERENCES:

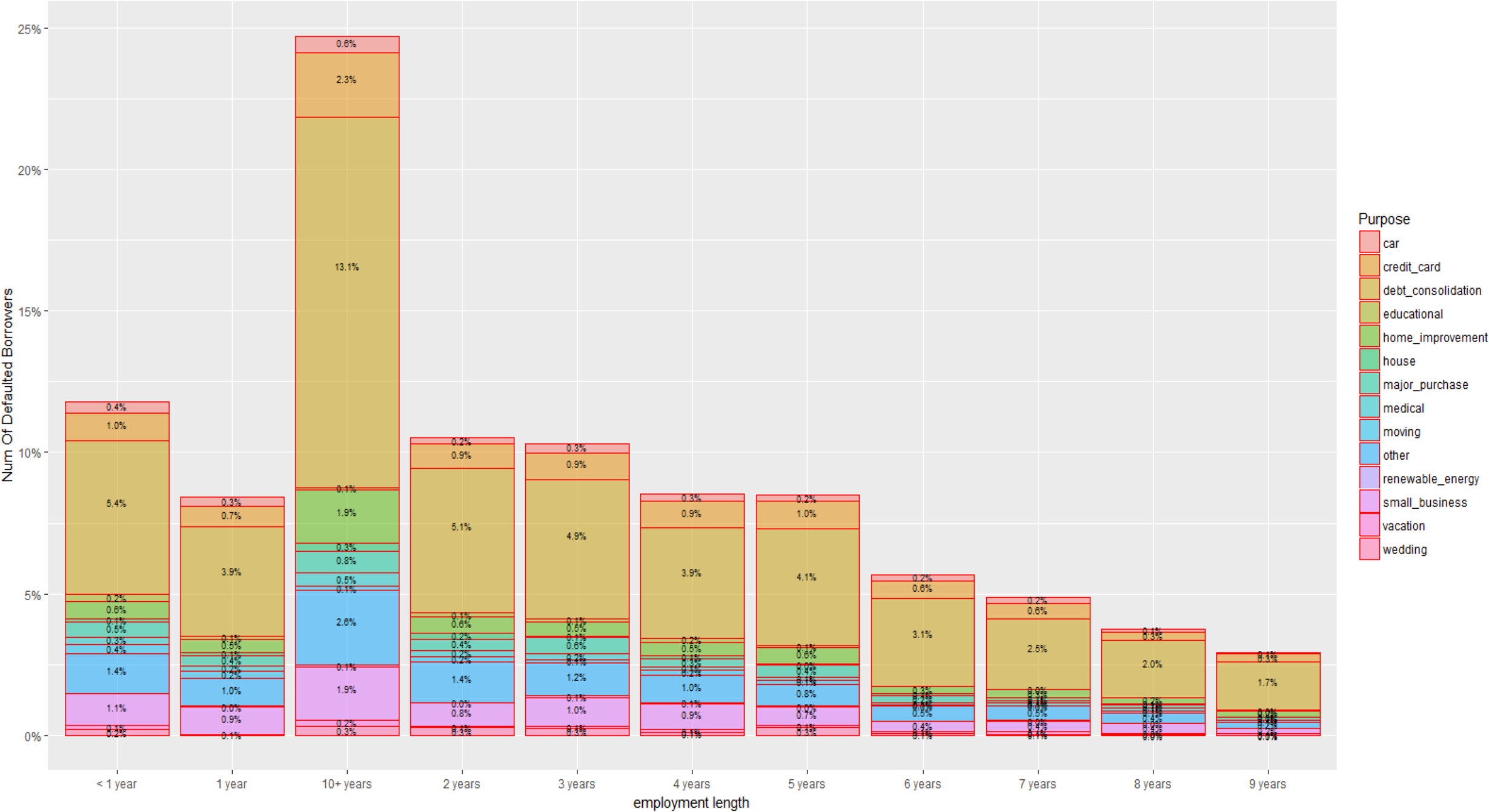
- Overall : 24.8 % of borrowers have 10 + years of employment
 - Out of them : 79.6 % of people have taken the loan in the interest range of 10 – 20%



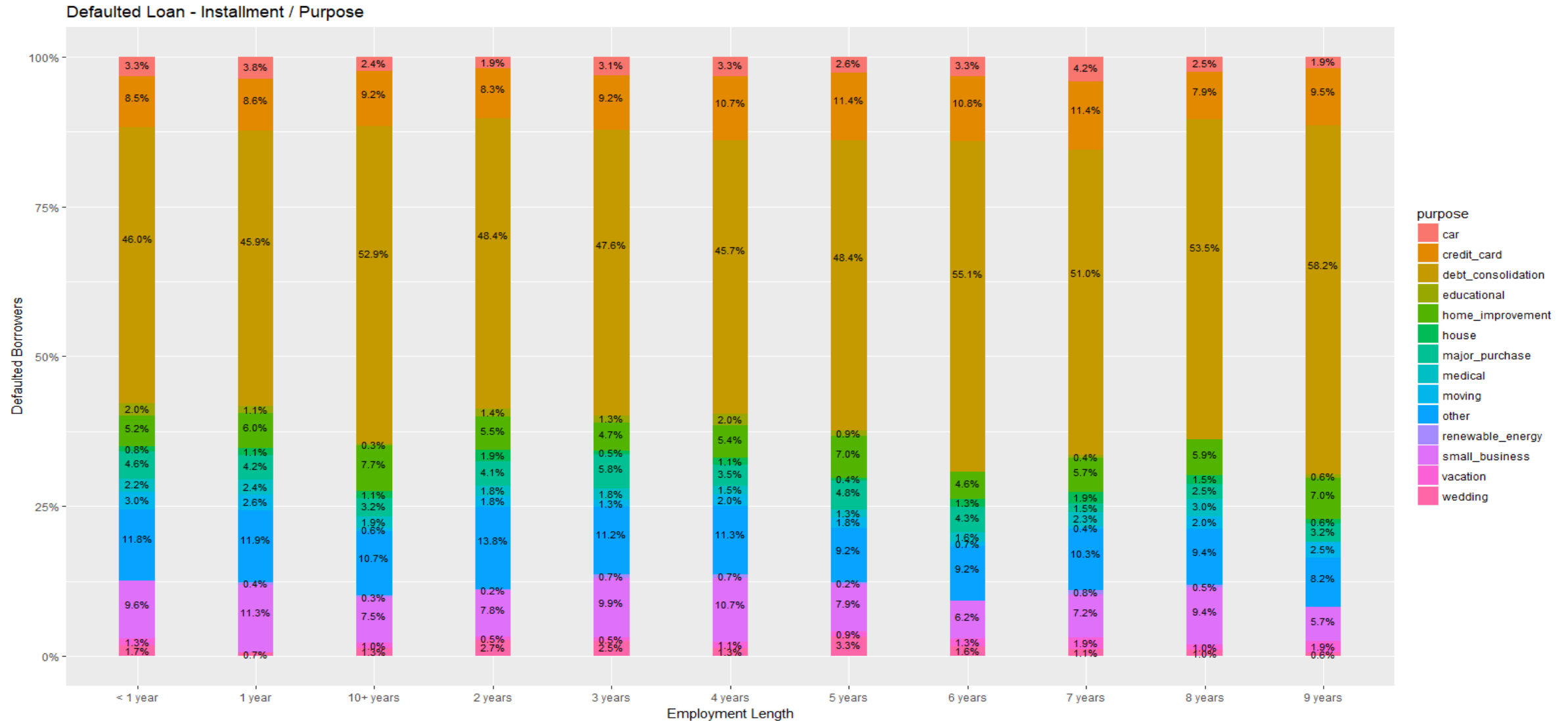
INFERENCES

- Grade B – loan defaulted : 25.2 % , Grade C : 100% , Grade D : 20.1 %
- Out of them
 - Grade B – Interest range (10 – 20 %) defaulted = 84.9 %
 - Grade C and Grade D - Interest range (10 – 20 %) defaulted = 100 %

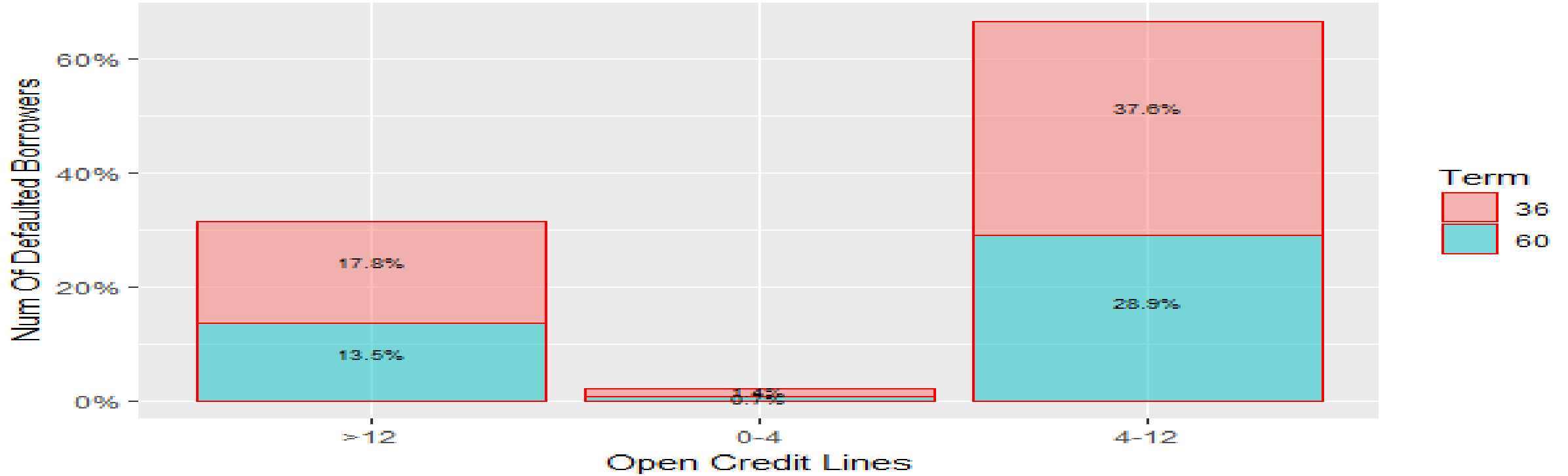
Defaulted Loan - Employment Length / Purpose



- Overall : (employment > 10 years) + (employment <= 1 year) = 45%
- Out of them : nearly 48 % have taken the loan for debt consolidation:



Defaulted Loan - Open Credit Lines / Term



- **INFERENCES**

- Overall 66.5 % of borrowers defaulted those have open credit lines in the range of 4 -12
 - Out of those : 56% of those loans have a term of 36 months

Overall INFERENCE

LOAN DEFAULT SCENARIOS:

1. First Scenario

- annual_inc_range -- MEDIUM i.e. 20k - 80k
- installment_range – MEDIUM i.e. 150 - 400
- funded_amnt --- LOW i.e. 0 - 10K
- int_rate ---- 10 - 20 %
- emp_length --- 10 + years
- purpose ---- debt consolidation

2. Second Scenario

- int_rate ---- 10 - 20 %
- emp_length --- 10 + years
- purpose ---- debt consolidation