LOAN RISK ANALYSIS CASE STUDY

SUBMISSION

Group Name:

1. Ashwin Rangarajan - DDA1730137
2. Prachi Prakash – DDA1730058
3. Shikha Chaturvedi – DDA1730288
4. Sumanth Sundaramurthy - DDA1730138

Business Objective – Recognise driving factors of charged off loans and utilize the knowledge in risk assessment

While working for a consumer finance company which specializes in providing loans for various loan products to urban customers, there could be two types of risks associated with the bank's decision on loans (which are approved based on the applicant's profile):

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

Business Strategy:

The Finance company wants to analyse historical loan data and identify trends and traits of customers who have charged off/defaulted on loans

- The strategy of the analysis is to perform:

- **Univariate analysis**: Analysing past customer loan applicant data of the finance company and identify which customers are best fit for loan approvals
- **Bivariate analysis**: Analysing interdependency of various loan applicant data and derive insights from the same

Problem Solving Methodology

Data Collection

- Loan data (loan.csv) has been provided as part of the case study
- Data dictionary (Data_Dictionary.xlsx) has also been provided for definition of fields in loan dataset
- Raw data was loaded into R Studio

Data Cleaning

- Necessary data cleaning techniques were followed such as:
  - Removal of fields with greater percentage of NA values
  - Setting up required fields as factors e.g. term, loan grade, sub-grade, employee title, home_ownership , verification_status, loan status, pymnt_plan, purpose, addr_state, initial_list_status, application_type
  - Formatting of date fields to derive required time dimensions – loan issue month, year

Data Analysis

- Univariate Analysis – The fields of the data set was individually analyzed to gain insights. Eg: Insights were made based on trends relating to loan amount, term, int rates, installment.
- Segmented Univariate Analysis – The popular loan products offered by the company were found out based on the purpose of the loan. The individual variables were analyzed for top five loan products
- Bivariate Analysis – While univariate analysis gives insights on understanding the data, bivariate analysis helps further to the cause of exploring the trends and indentifying patterns

Data Visualization

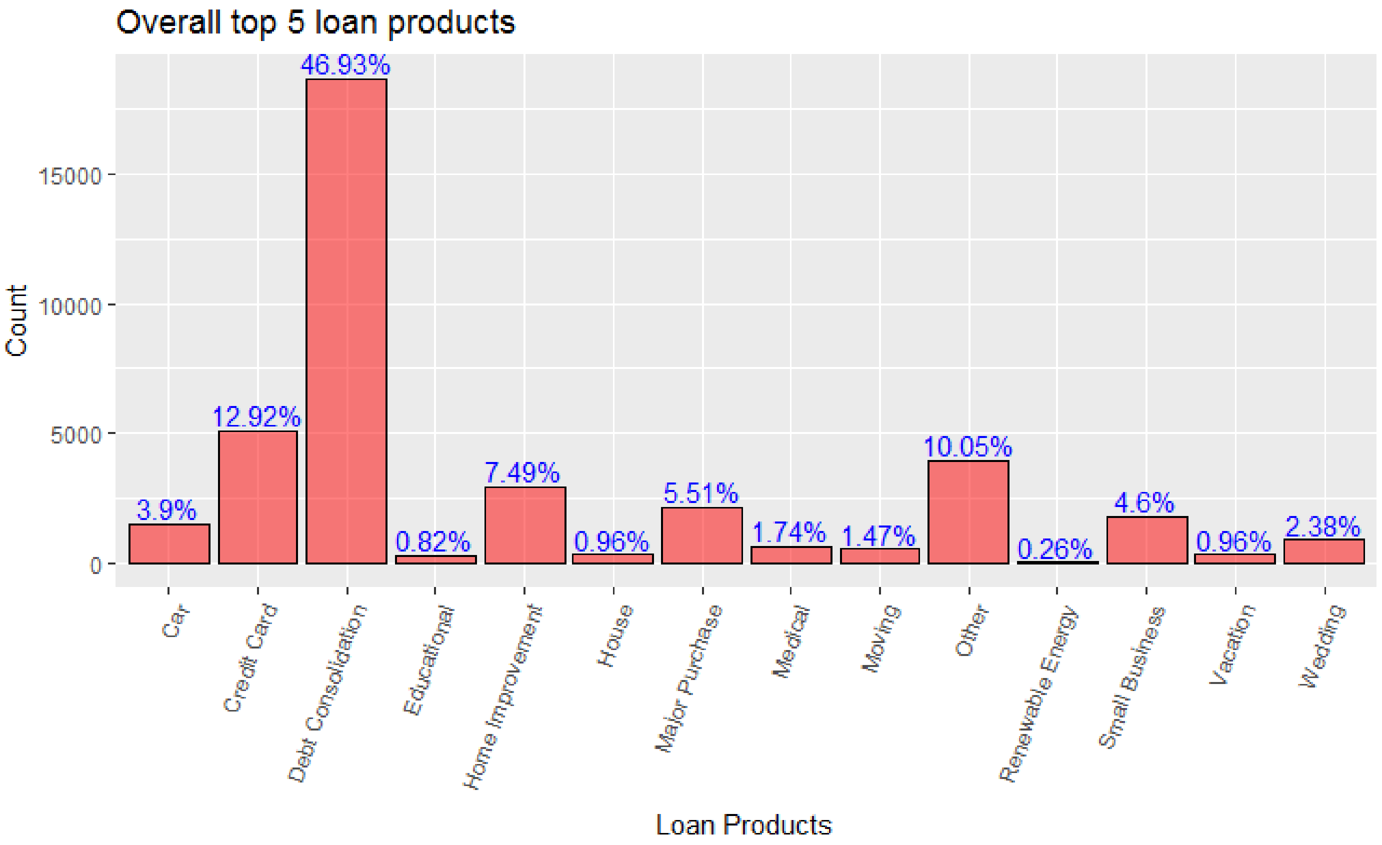
- Easiest way to find trends in the data is to plot the variables of interest
- Plots such as bar plots, histograms, density plots, box plots, stacked bar charts and so on are made using R tool to gain useful insights

Data Analysis

Variable of Interest – **loan_status = "Charged Off"**

Important Driver Variables	Important Driver Variables
loan_amnt – Done	annual_inc – Done
Term – Done	int_rates – Ashwin to update
Installment – Done	open_acc – Done
emp_length – Done	revol_util – Done
home_ownership – Done	Dti – Done
inq_last_6mths – Done	verification_status – Done
Purpose – Done	Grade and Subgrade – Ashwin to check

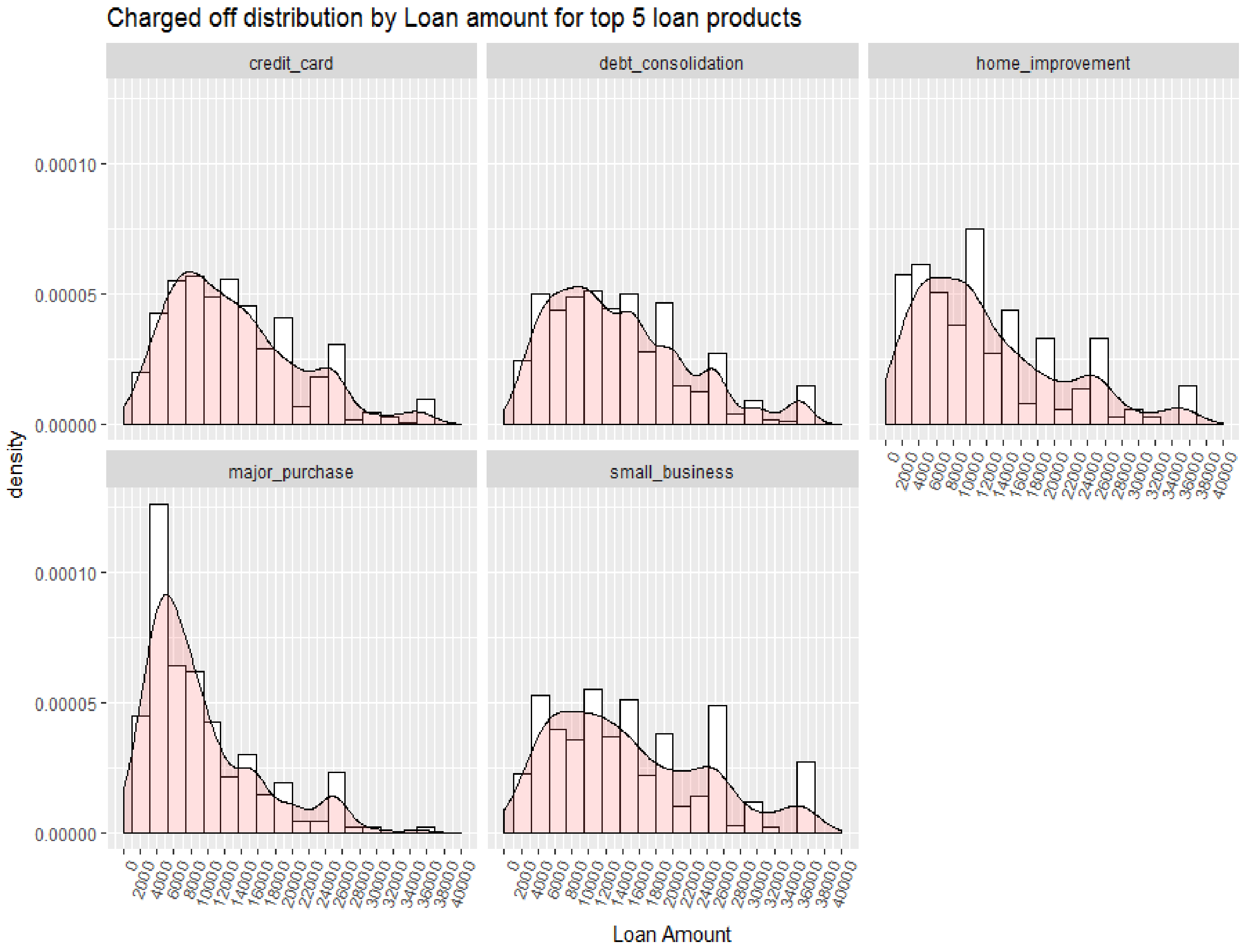
Distribution of Loans by Loan purpose



Observations –

- Five popular loan products are : debt_consolidation, credit_card, home_improvement, major_purchase, small_business

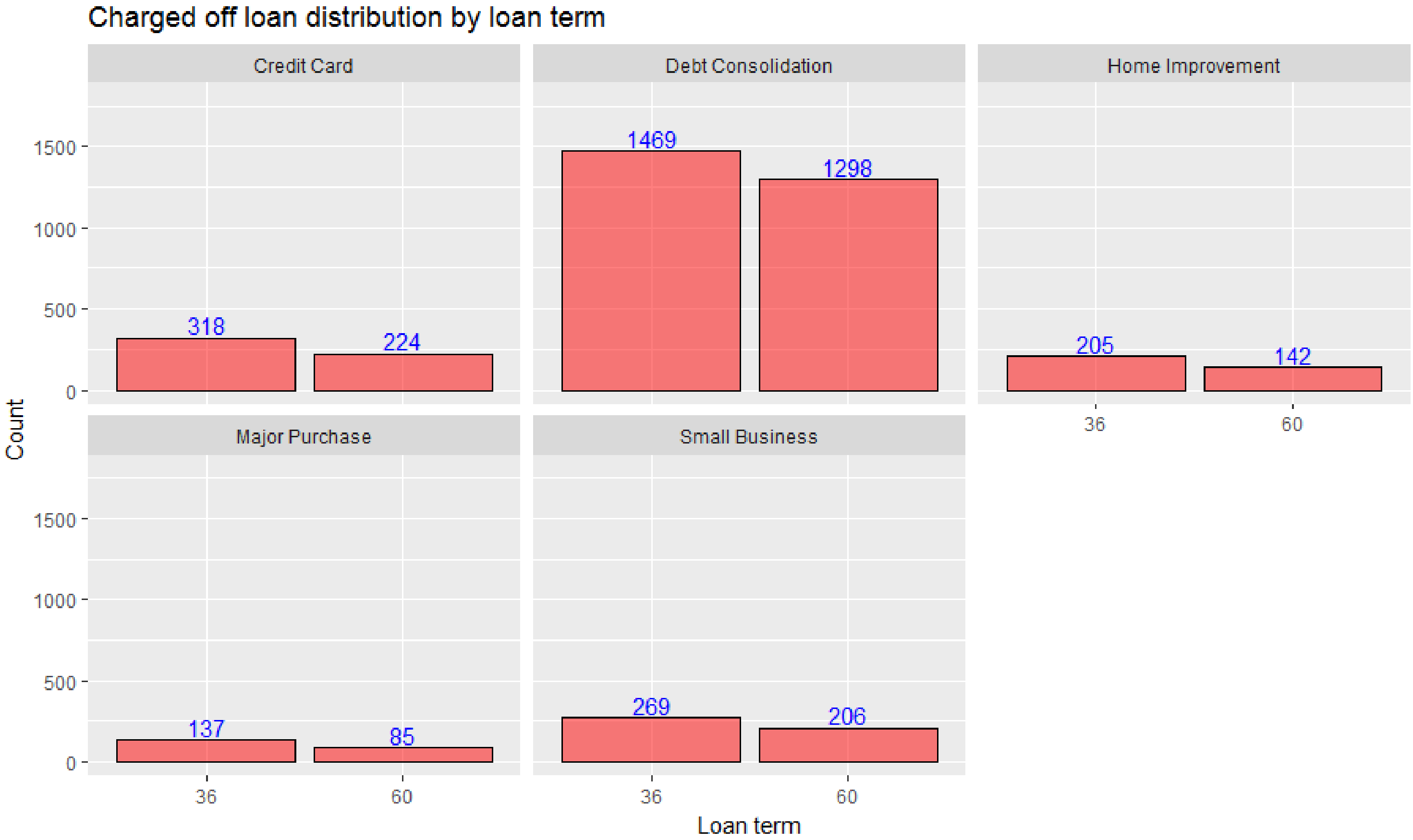
Distribution of Charged off loans by Loan amount for top 5 loan products



Observations –

- It is the amount of loan applied by the borrower; The applicants taking loan amount ranging from 3000 to 5000 are likely to default
- The graph is right skewed for almost all the loan products. But loan amount taken for Major Purchase and defaulting have high density

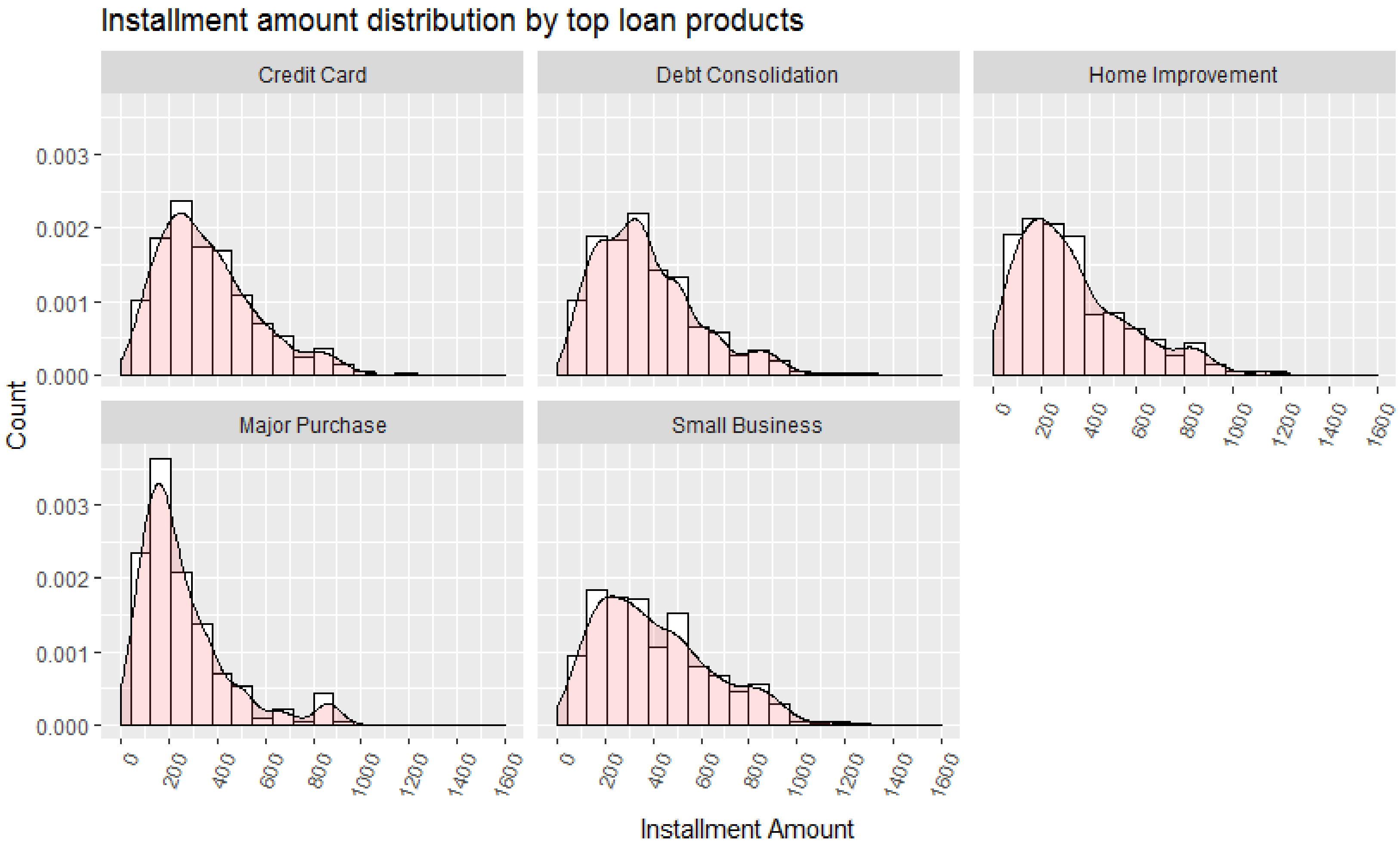
Distribution of Charged off loans by Loan term for top 5 loan products



Observations –

- 61.25% applicants have defaulted with 36 months of term for debt_consolidation loan product whereas 66.39% applicants have defaulted for 60 months of term for the same product

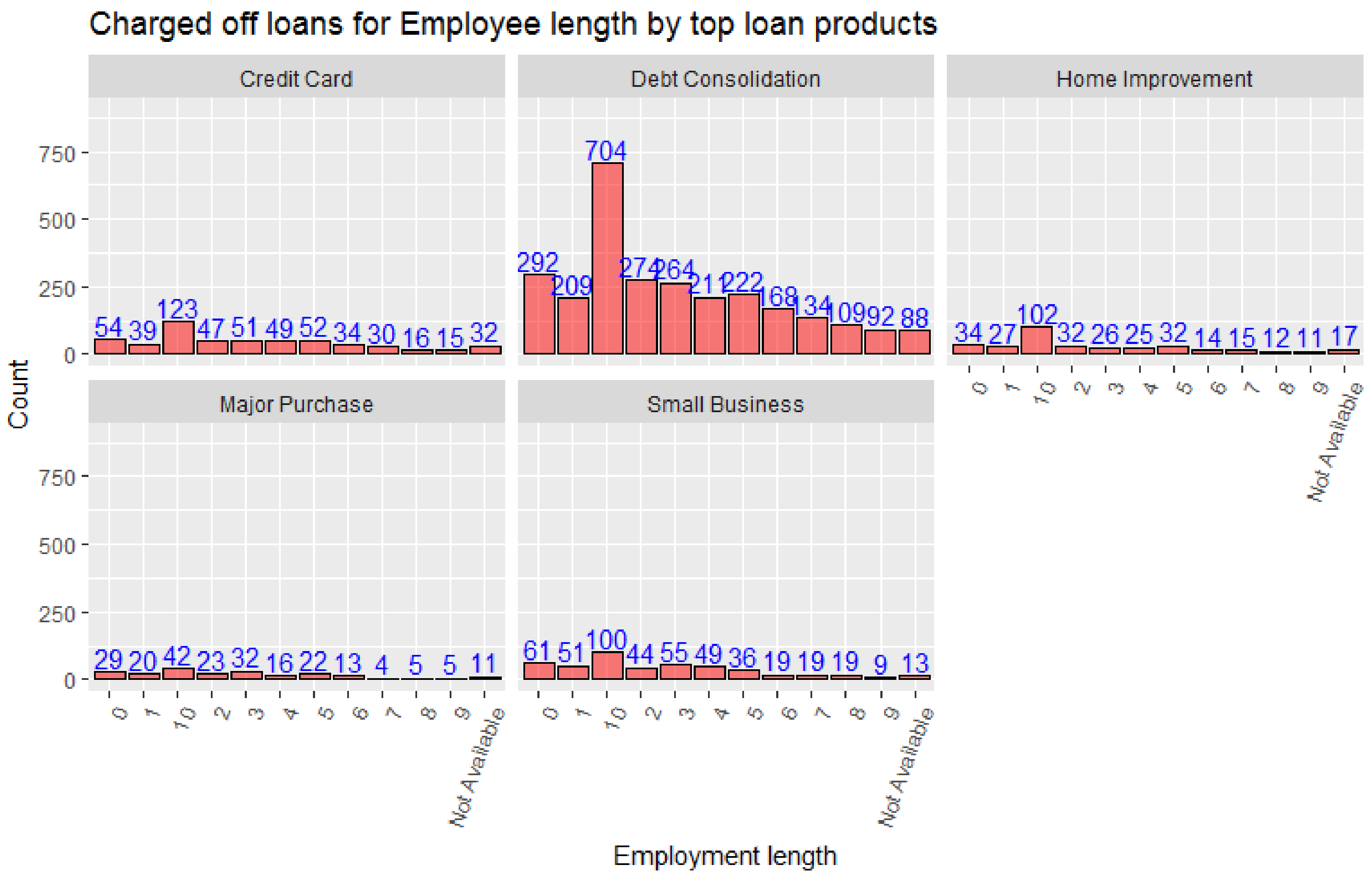
Distribution of Charged off loans by Loan Instalment for top 5 loan products



Observations –

- The applicants taking loan for **major purchase** have high density around 0-200

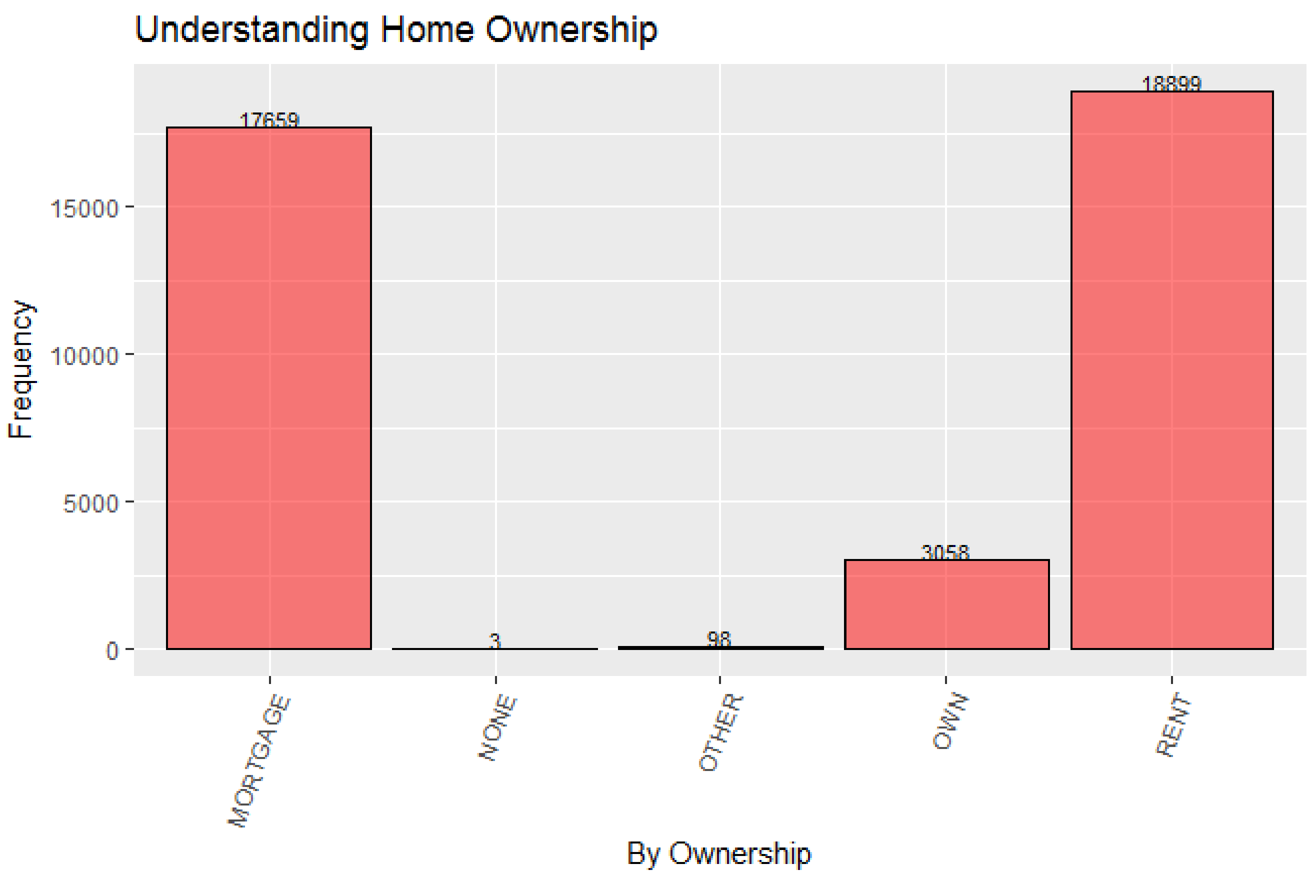
Distribution of Charged off loans by Employee length for top 5 loan products



Observations –

- The applicants having 10 or more than 10 years if employment are likely to default.
- Additionally, there are high number of defaulters who have taken loan for **debt consolidation**

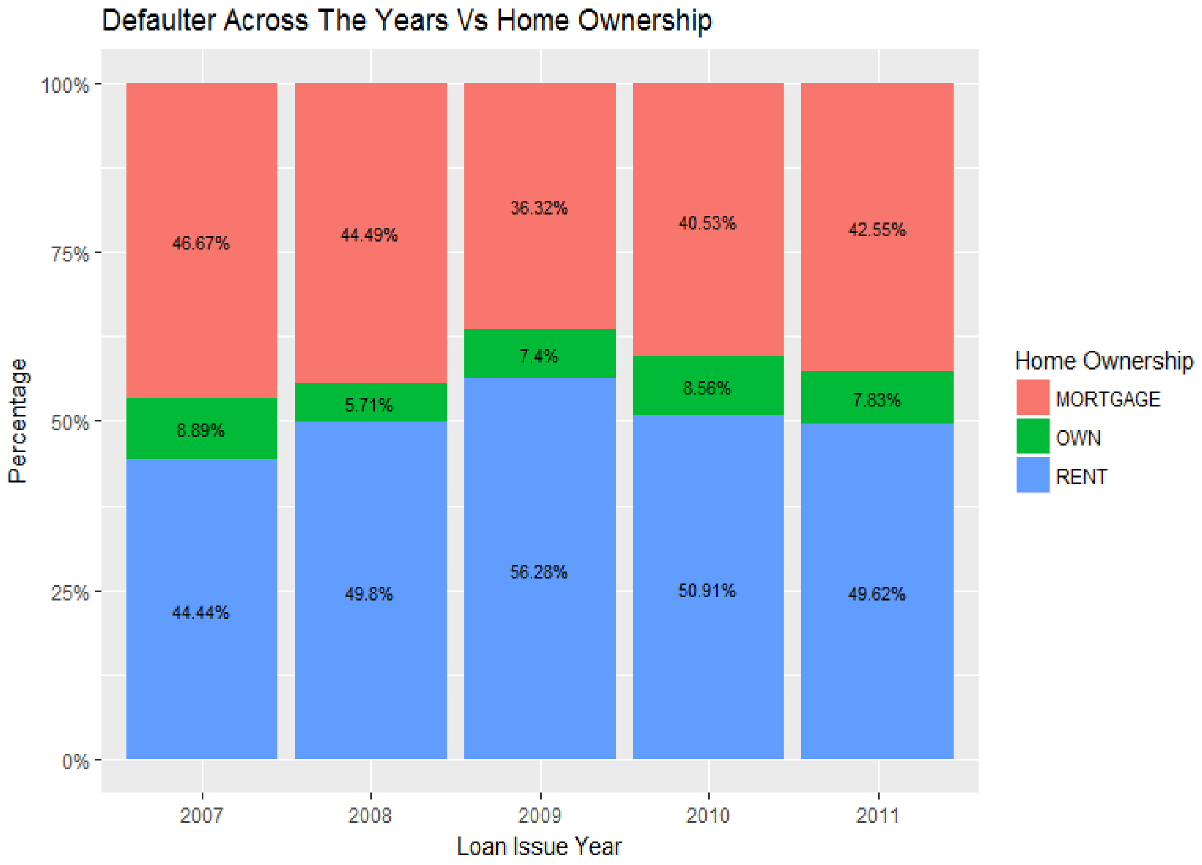
Distribution of Loans by Home ownership



Observations –

- From the plot one can see that the more applicants RENT a house rather than OWN/Mortgage

Distribution of charged off loans by Home Ownership across years and loan products



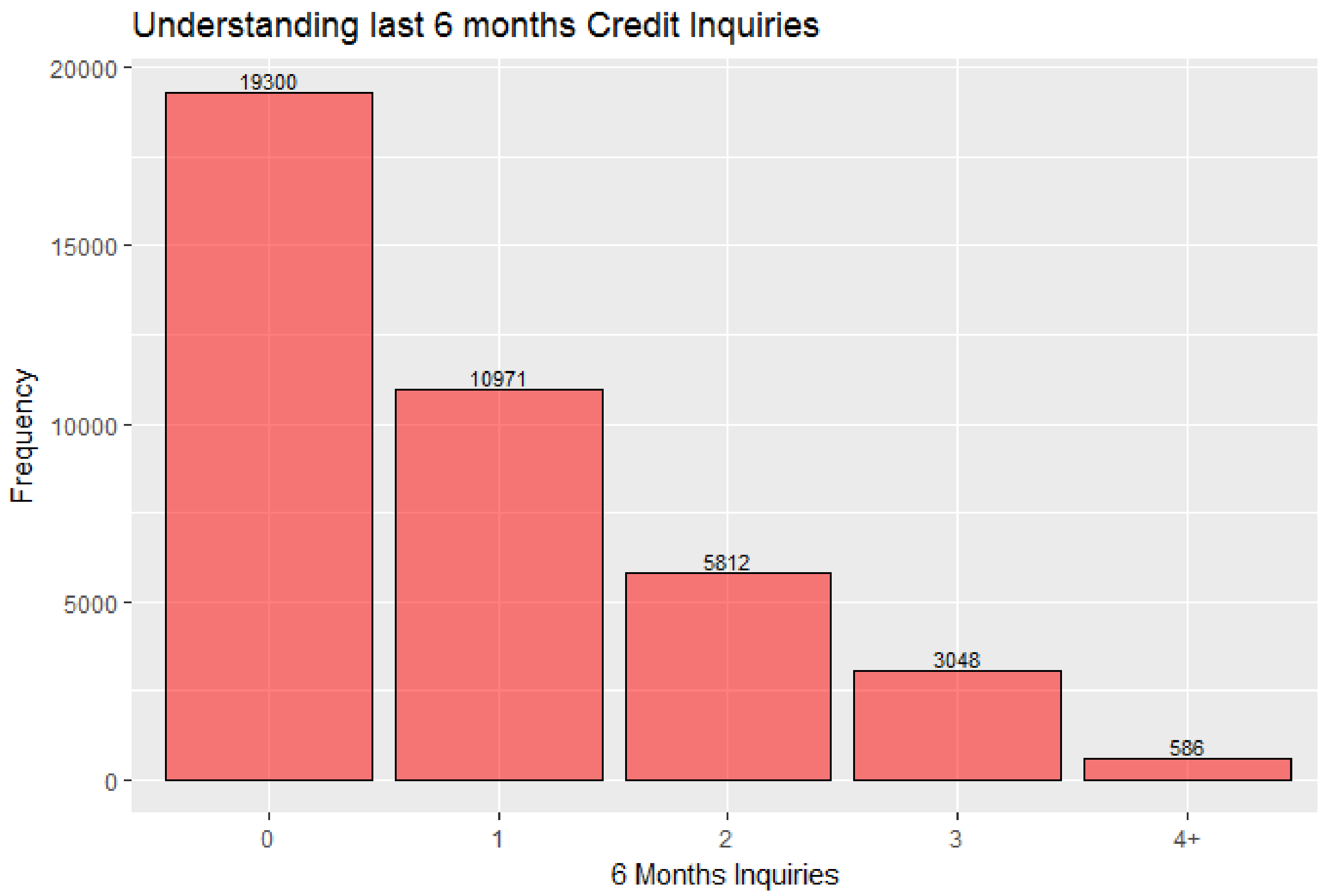
Observations –

- Loan products except 'Home Improvement' follows the same trend where applicants who have rented the house are likely to default.

Observations –

- Applicants who rent a house are more likely to default

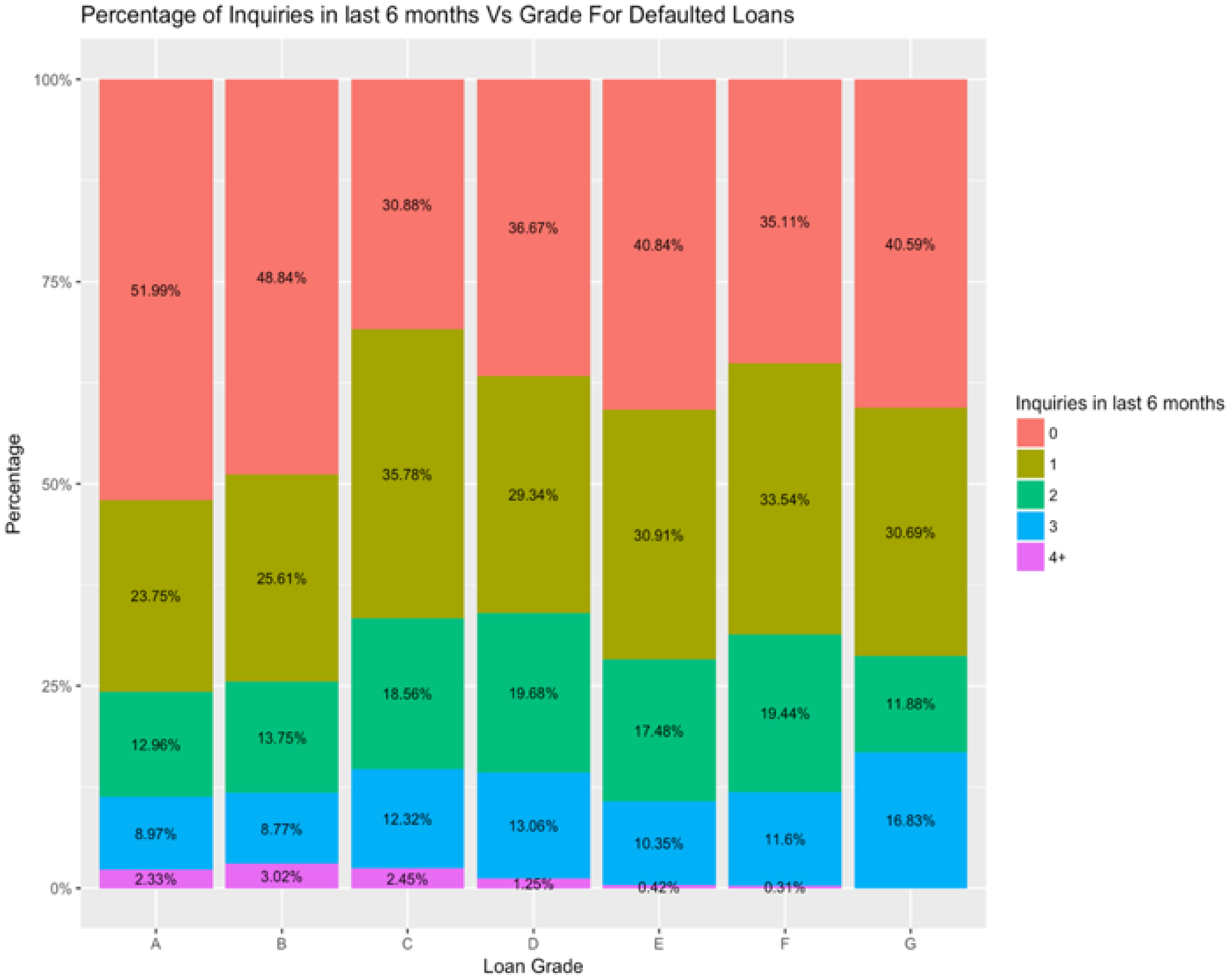
Distribution of Loans by Credit enquiries in last 6 months for top 5 loan products



Observations –

- directly using the credit inquiries will not reveal much however lenders with very high number of inquiries are considered as risky applicants as it is an evidence for more credit seeking behavior.
- In the entire data set with all loan status folks with 2 and above inquiries can be identified as risky applicants as they have higher inquiries within a span of 6 months.

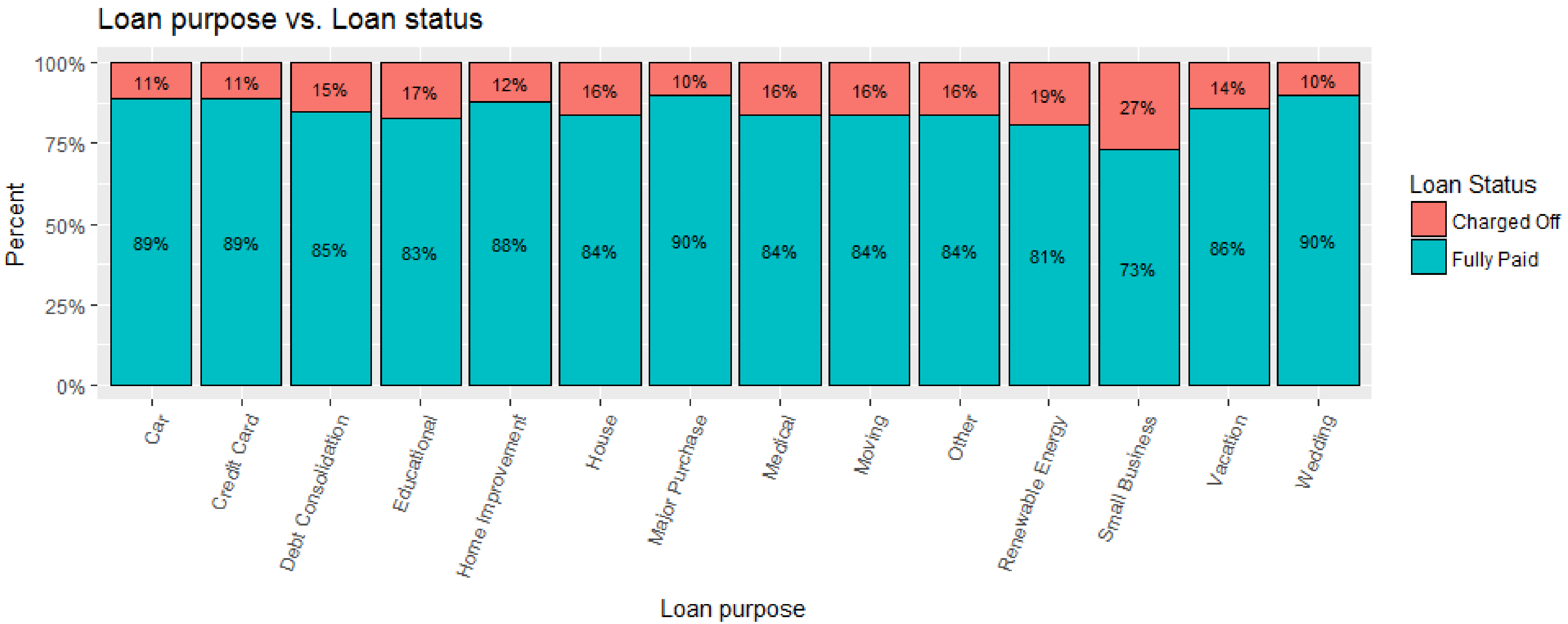
Distribution of charged off loans by number of Inquiries and Grade



Observations –

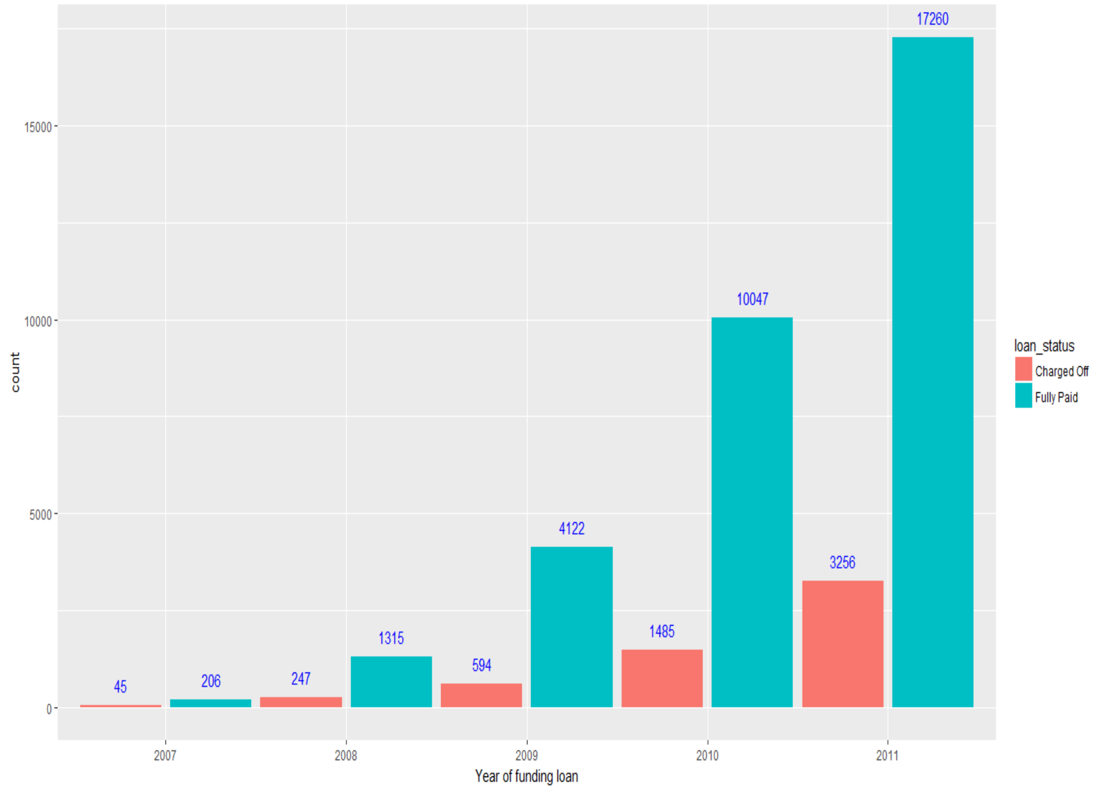
- Applicants with zero inquiries across different loan grades are risky clients because of their unknown credit seeking behavior. Data shows that majority of folks with 0 inquiries have defaulted more. It is better not to give them loan to avoid defaults or fund them a lesser loan amount.
- Applicants with 1 inquiries are also risky because with one inquiry on their credit file doesn't reflect their true credit seeking behavior. It is better to be cautious while giving loans to such applicants.
- Percentage of defaulters with 2 inquiries are relatively same across loan grades so they are less riskier applicants.
- Folks with 3 inquiries and above can also be classified as riskier applicants because within a short span to time they have a requirement of more credit and by not giving them loans one can reduce the percentage of default loans.

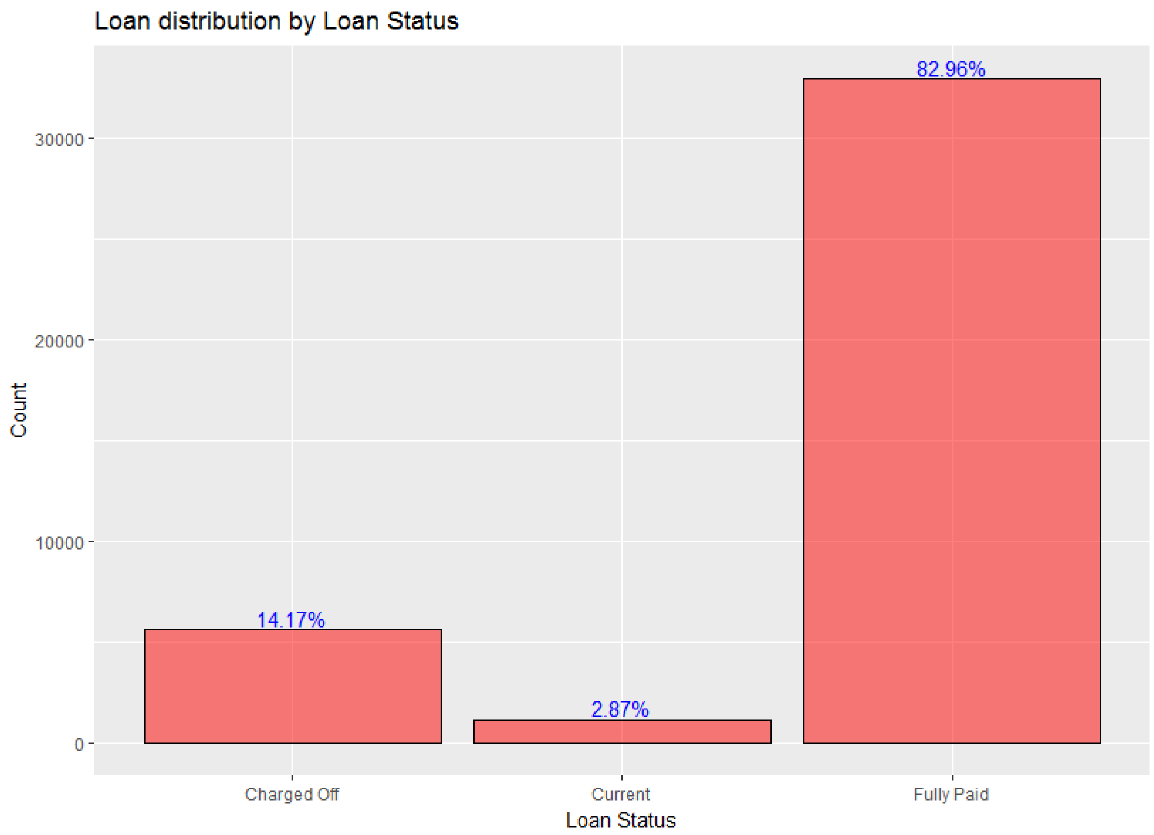
Top five loan products offered by the company and the variation across loan status



Top 5 loan products offered by the company is Debt Consolidation, Credit Card, Home Improvement, Major Purchase and Small Business

Distribution of Application count over the years and their default rate





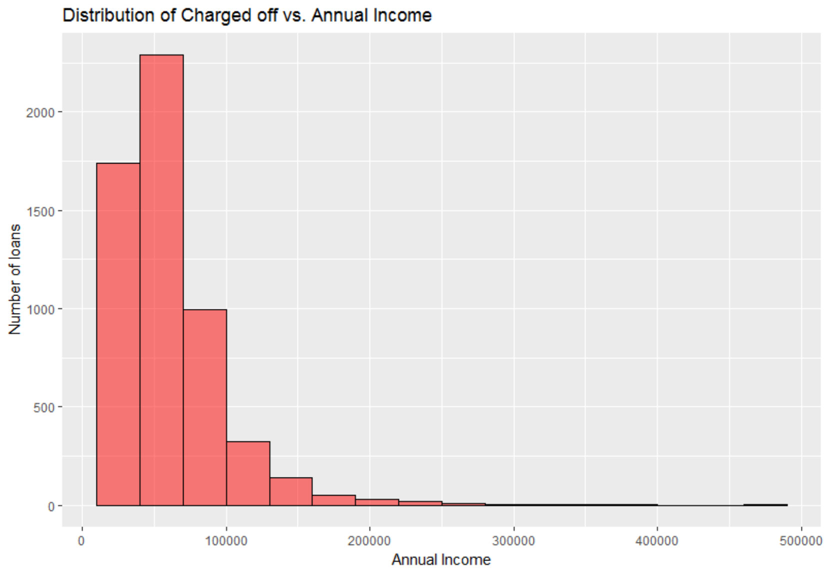
Observations –

- As the number of applicants have increased over the years from 2007 to 2011, there have been gradual increase in defaulters rate also over the same period of time

Observations –

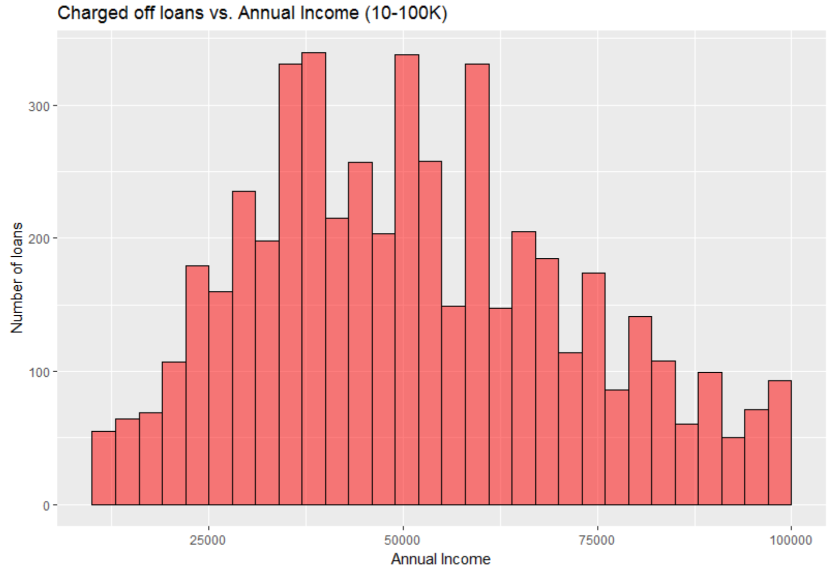
- We see that out of 39,717 applicants, 5,627 applicants have defaulted. The defaulter rate for the company is 14.17%

Distribution of charged off loans by Annual income



Observations –

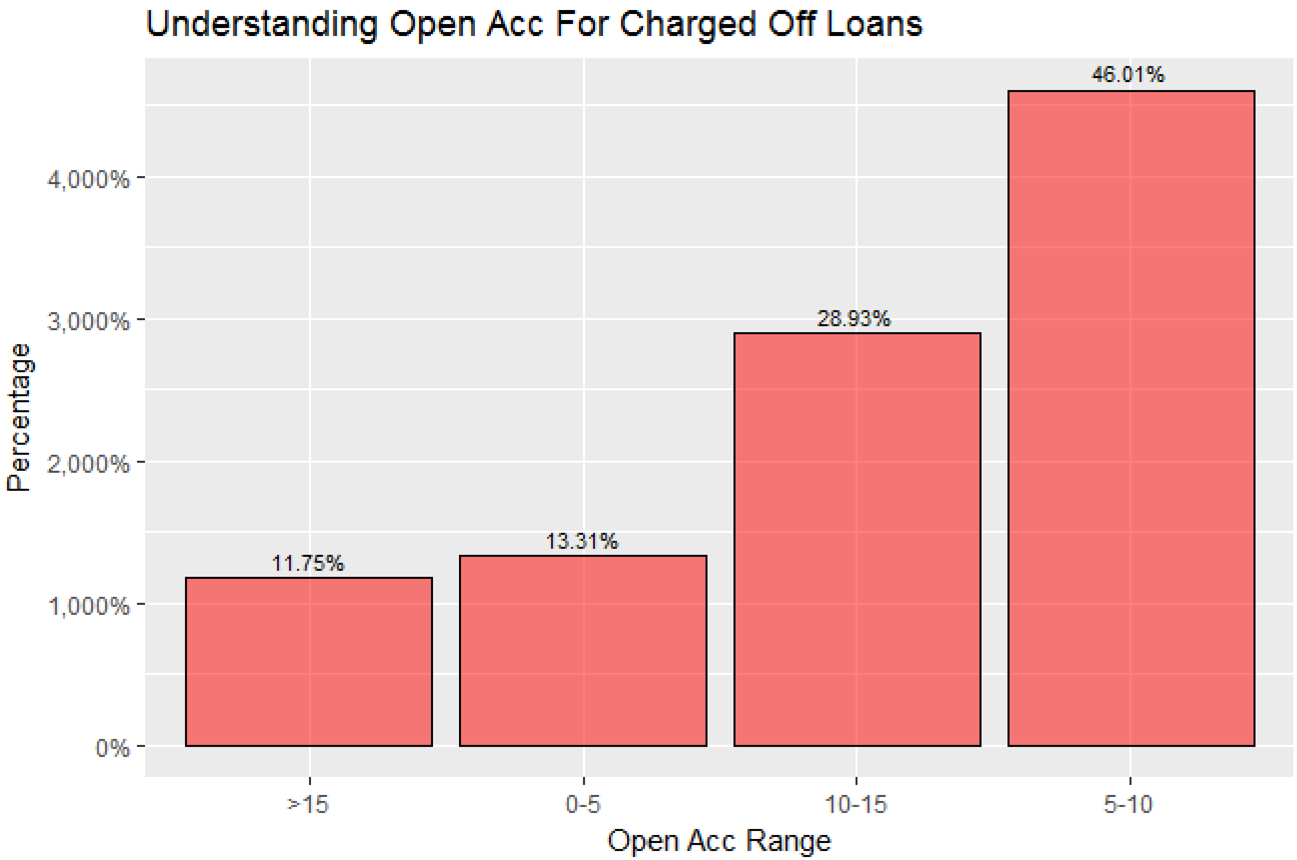
- We see the vast majority of the Charged off loans are by customers with Annual income < 100,000 USD



Observations –

- With a limited Annual income range (10-100K) and more granular bins we see that there is a higher majority of the Charged off loans by customers with Annual income between 25,000 to 75,000 USD

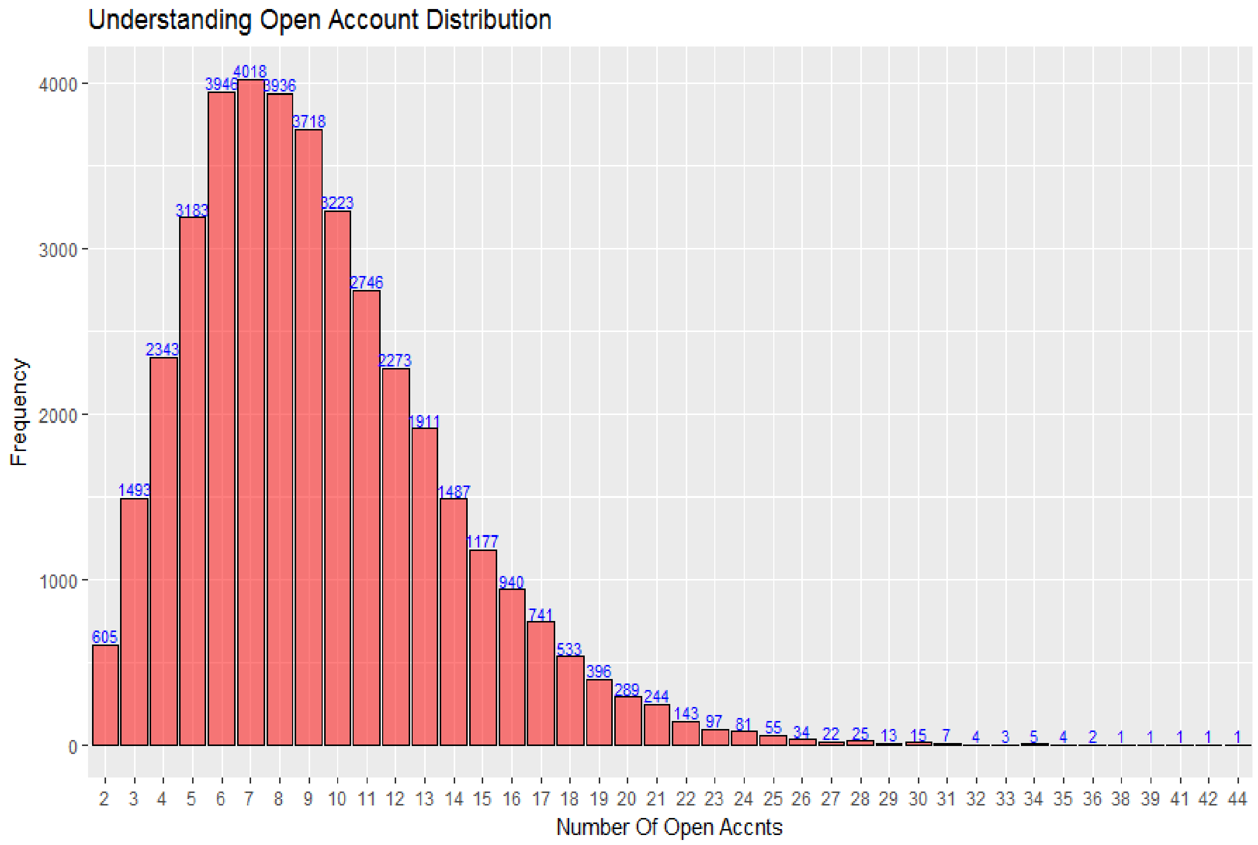
Distribution of Loans by number of open accounts

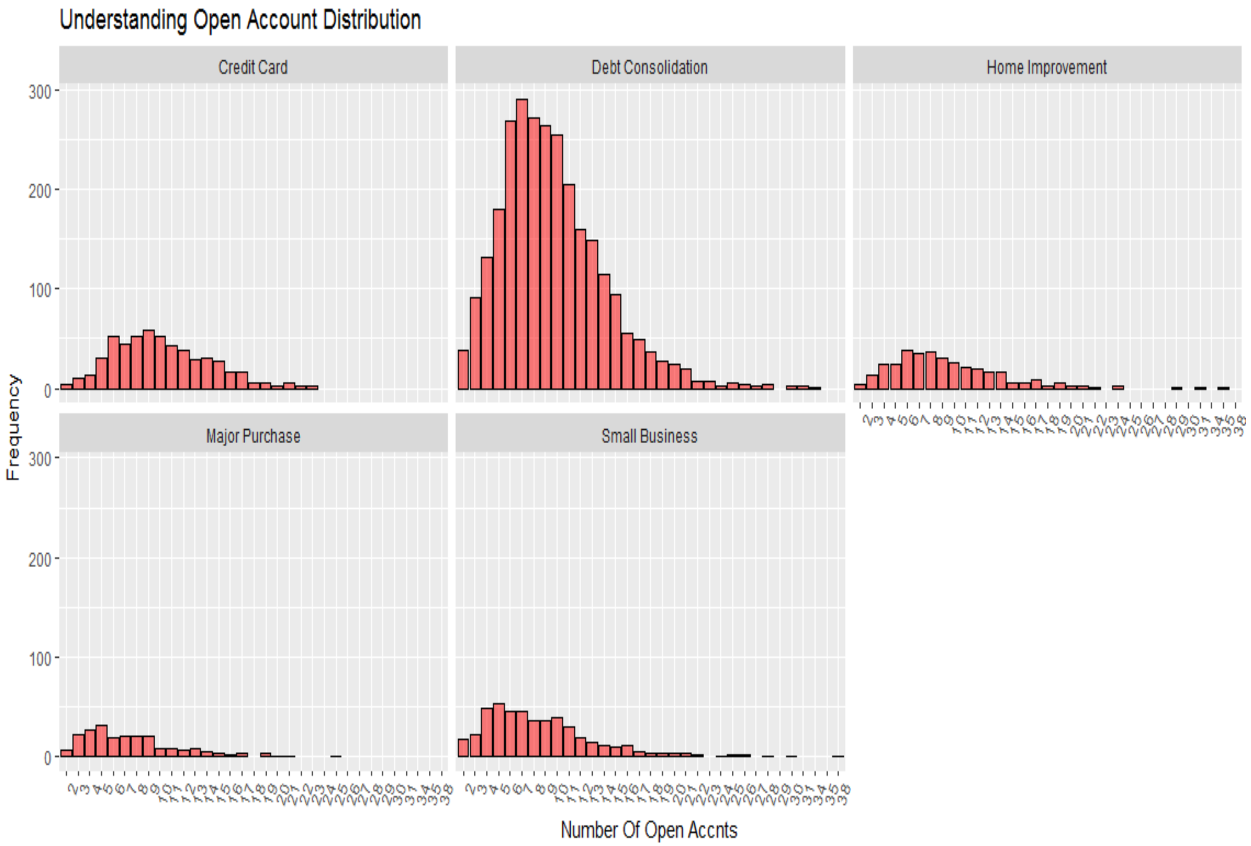


Observations –

- With the increase in number of Open Account beyond 5 the percentage of defaulting a loan increases drastically.
- It is advisable to give loans to applicants with open accounts up to 5 and reject the loan applications with open accounts above 5

Distribution of Loans by number of open accounts across loan products





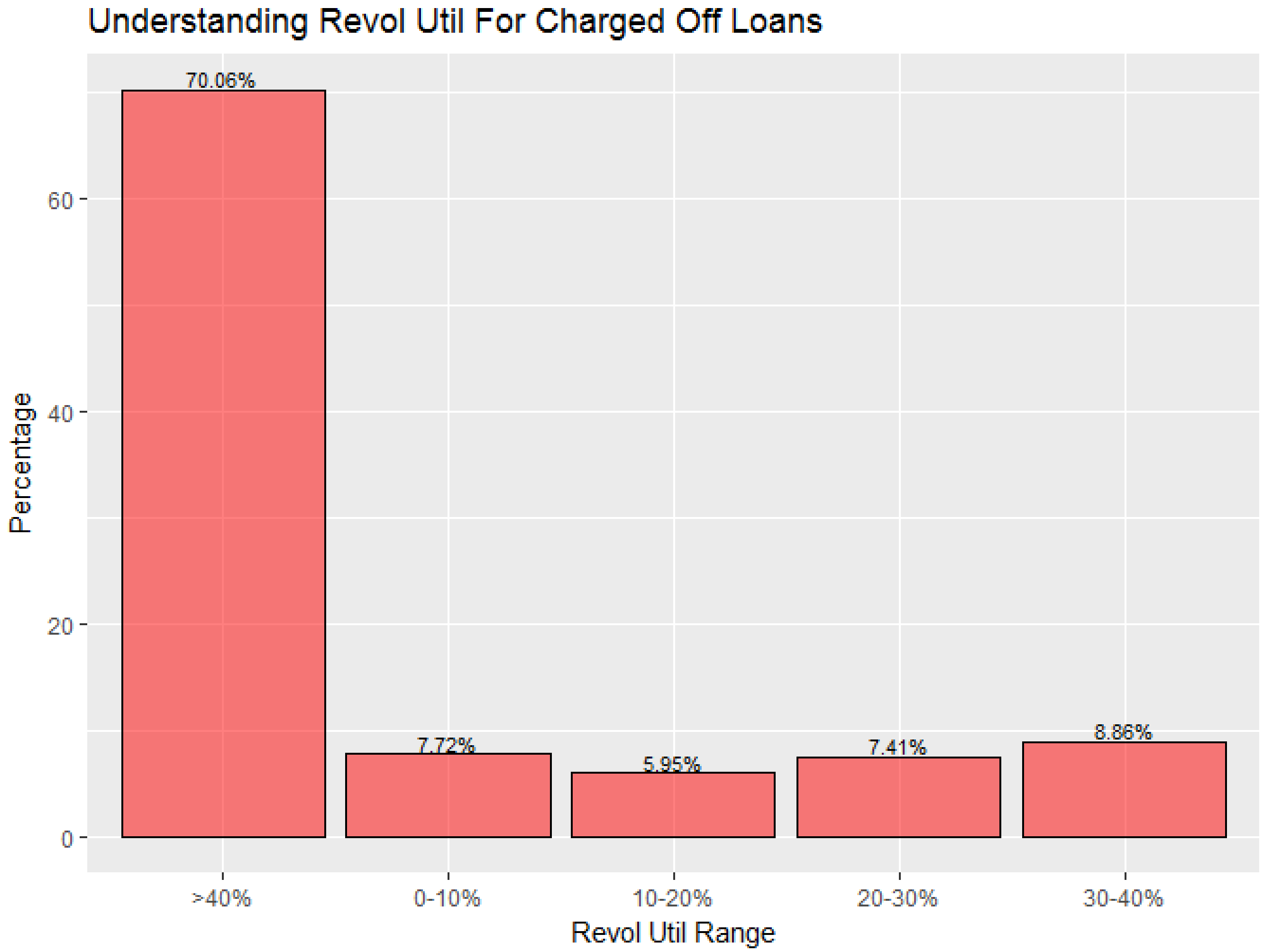
Observations –

- From the trend we can see that a minimum of 2 open accounts is necessary for getting a loan with the company. Overall the distribution of maximum loans have open account b/w 3 and 13

Observations –

- The graph across all product type is right skewed.

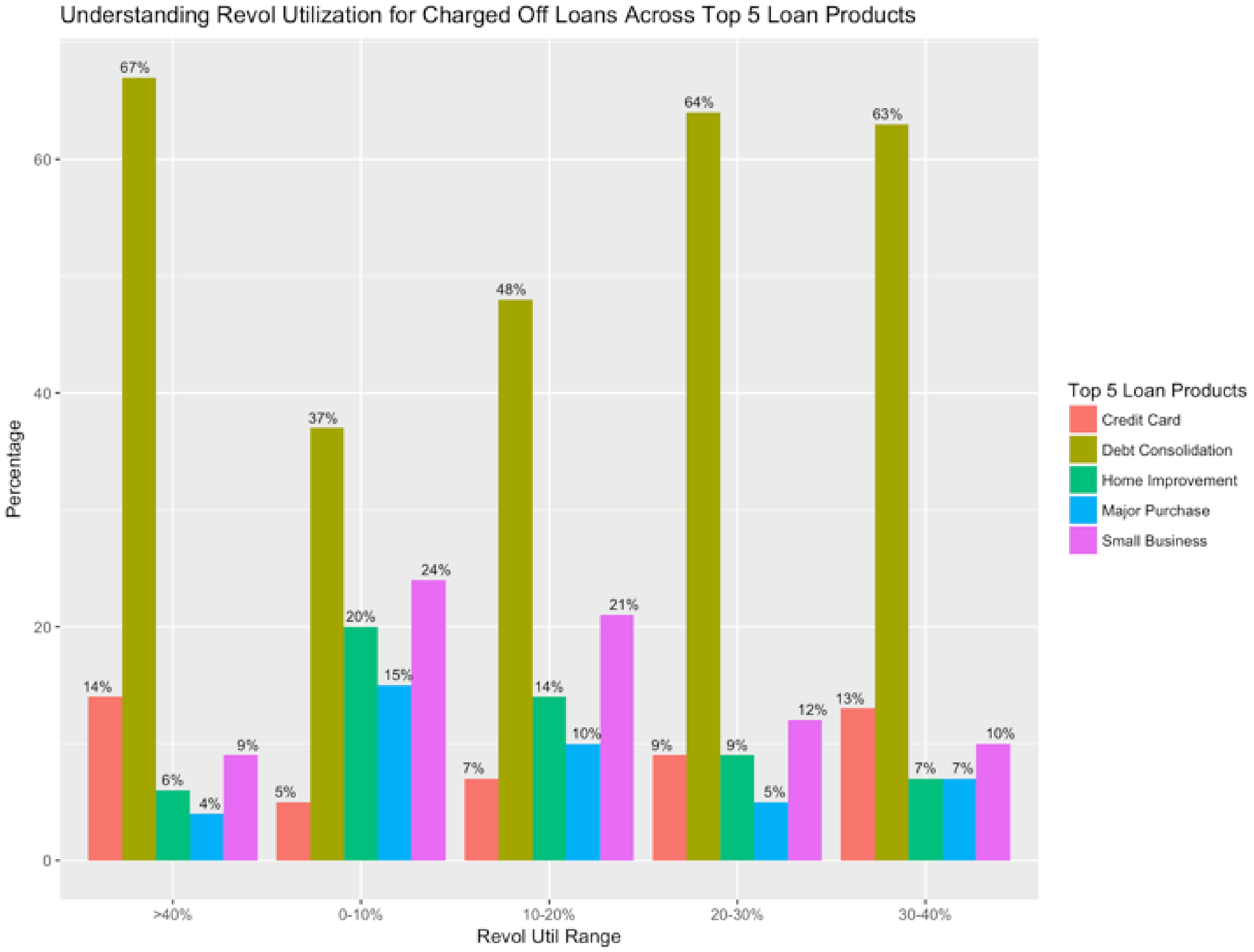
Distribution of charged off loans by Revol util range



Observations –

- The overall percentage of applicants with revolving utilization greater than 20% who have defaulted loans is about 86%. Such applicants are likely to default on their loans as they are not able to manage their credit line and are risky loan applicants.

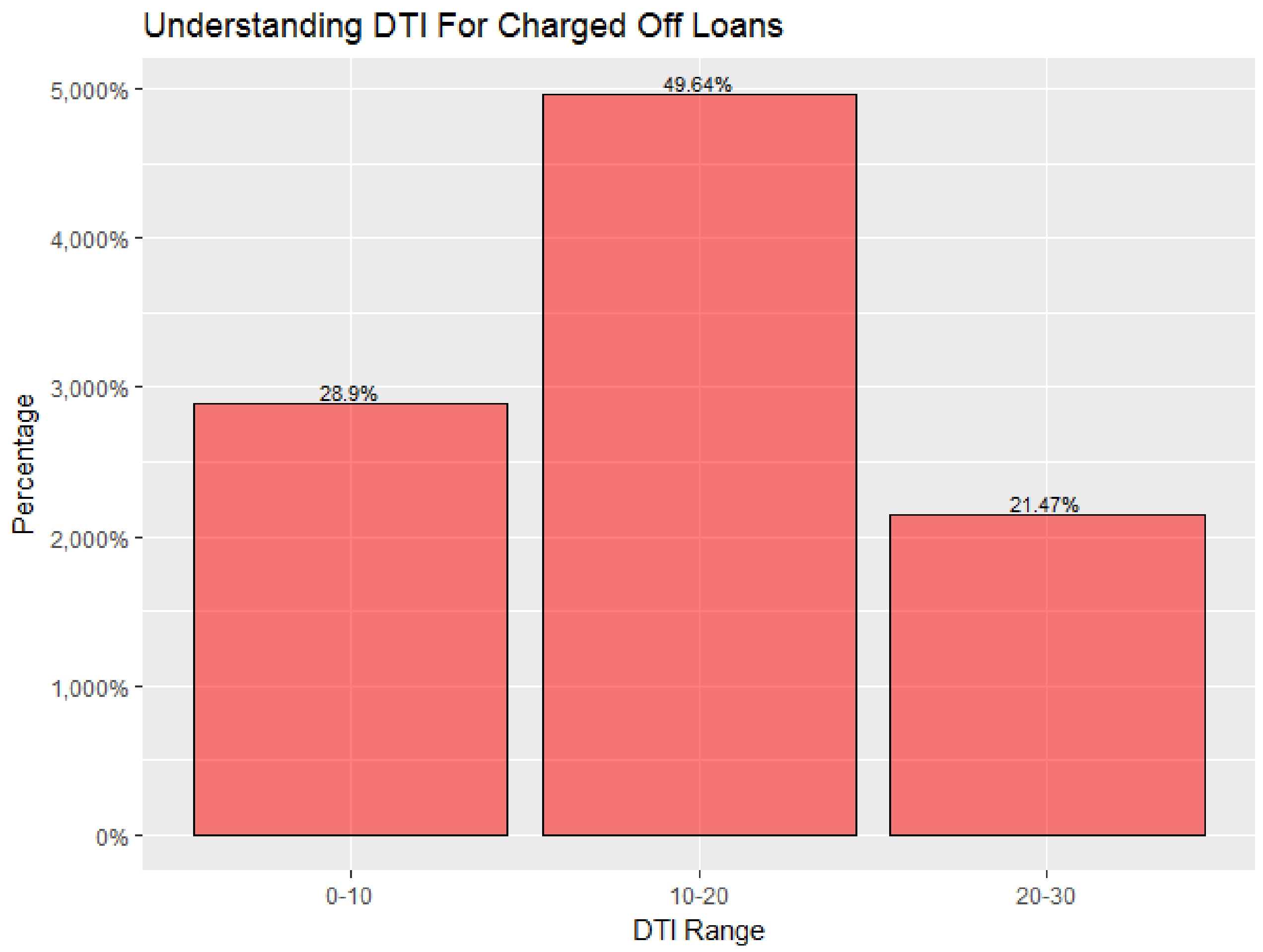
Distribution of charged off loans by Revol util range across loan products



Observations –

- For Debt Consolidation loan purpose, there is a drastic increase in loan defaults where the revolving utilization is greater than 20%.
- Applicants with revolving utilization greater than 20% especially for Debt Consolidation can be deemed as risky and likely to default on their loans.

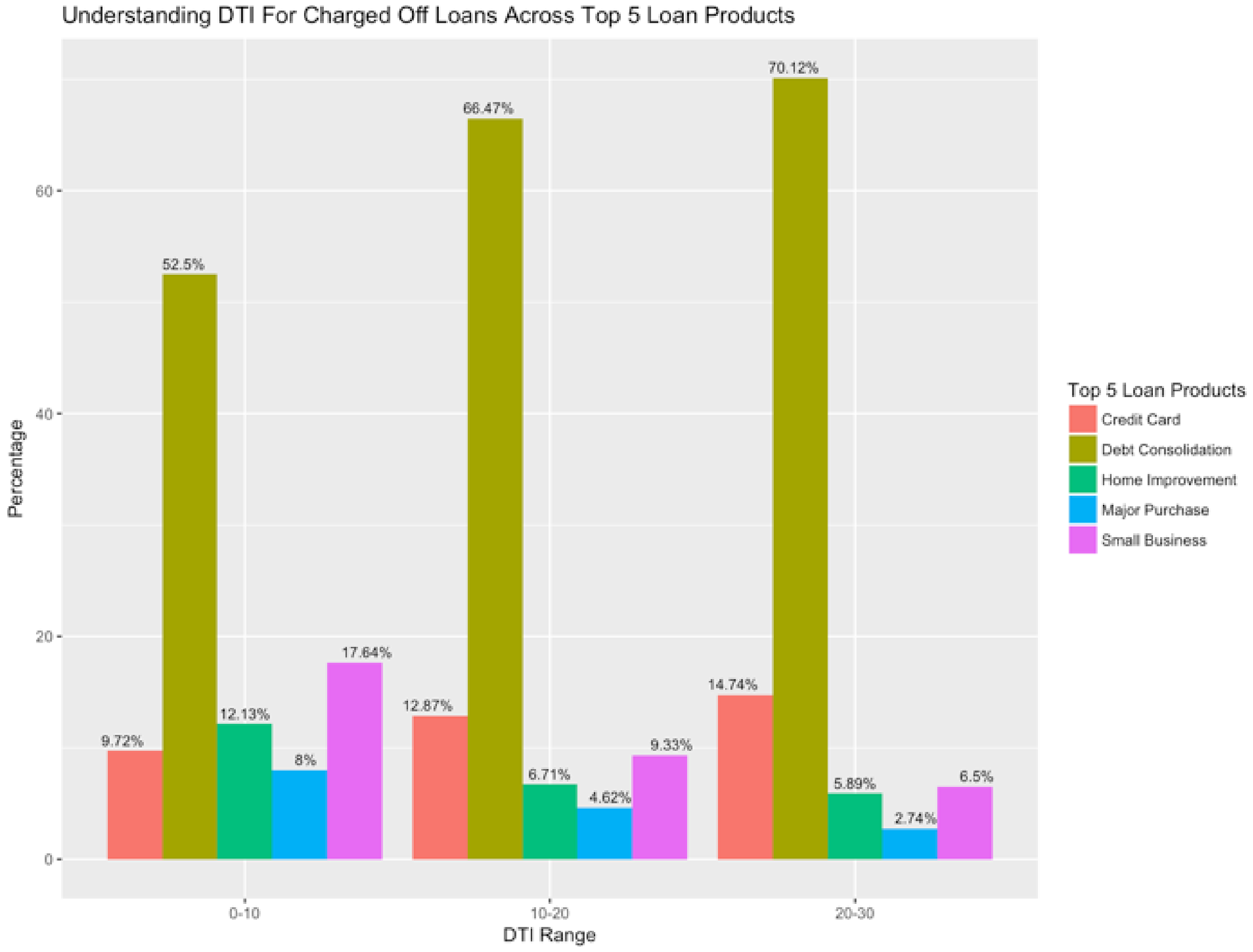
Distribution of charged off loans by DTI



Observations –

- As the DTI ratio goes above 10 the percentage of defaulting the loan also increases. If the DTI is available during the loan application processing then in order to avoid loan defaults, loans for applicants with DTI greater than 10 can be rejected or the funded amount can be reduced.

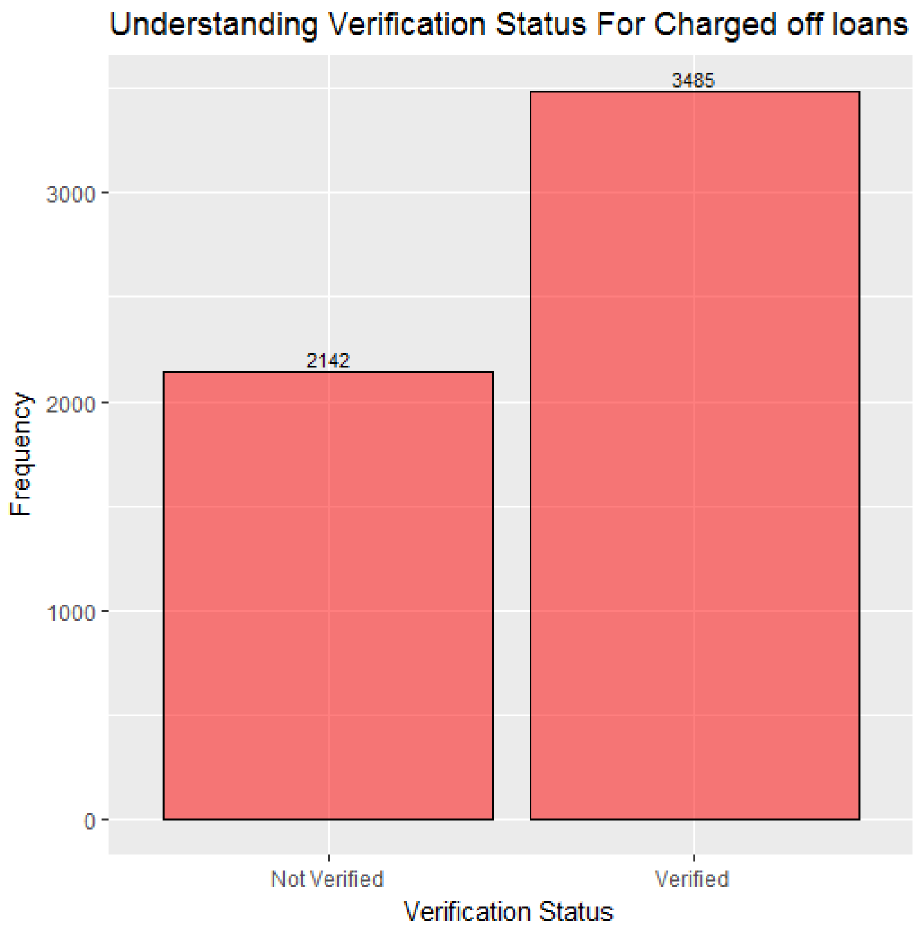
Distribution of charged off loans by DTI across all loan products

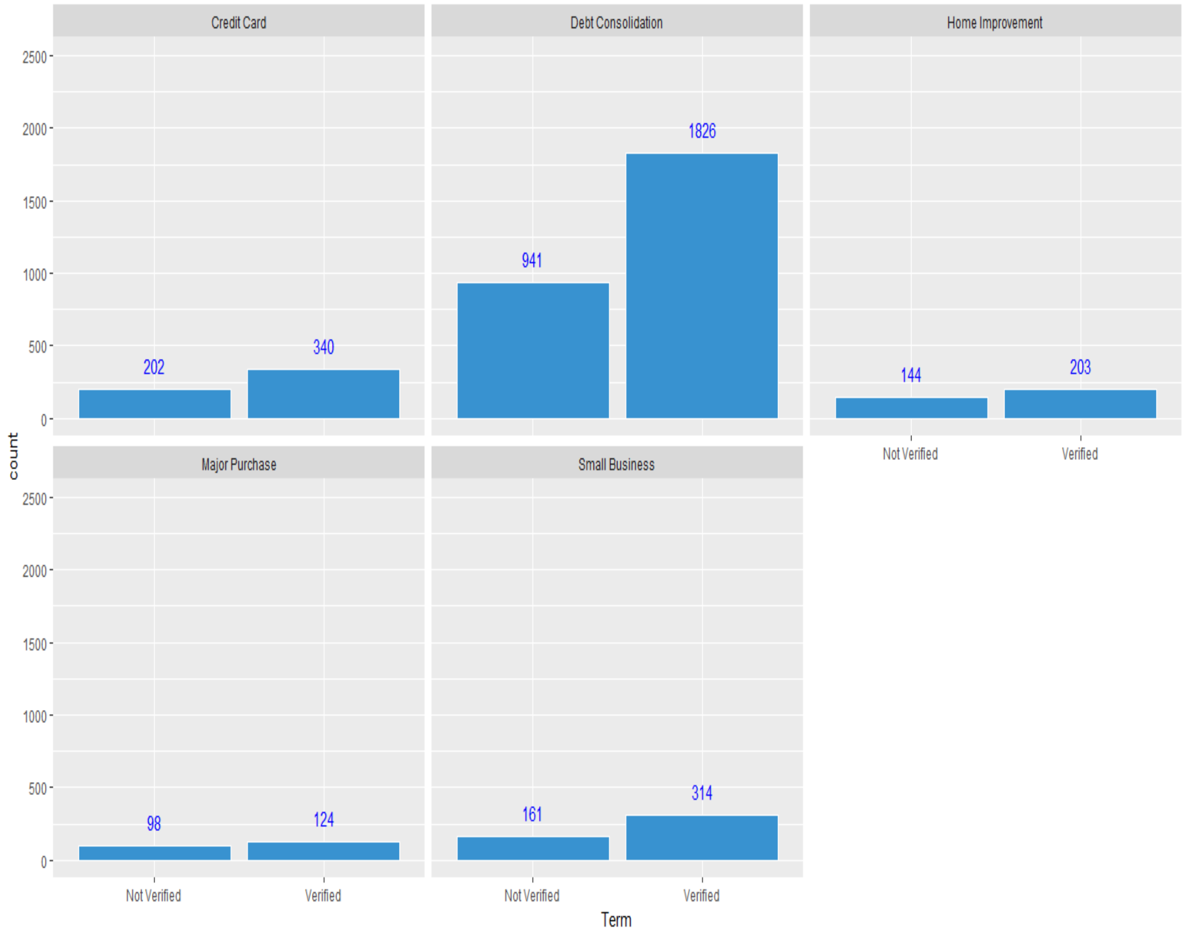


Observations –

- Across the top 5 loan product, debt consolidation with DTI greater than or equal to 10 seems to have higher chances of defaulting. Applicants with DTI >= 10 and loan purpose of DEBT Consolidation are likely to default.

Distribution of charged off loans by Income verification status





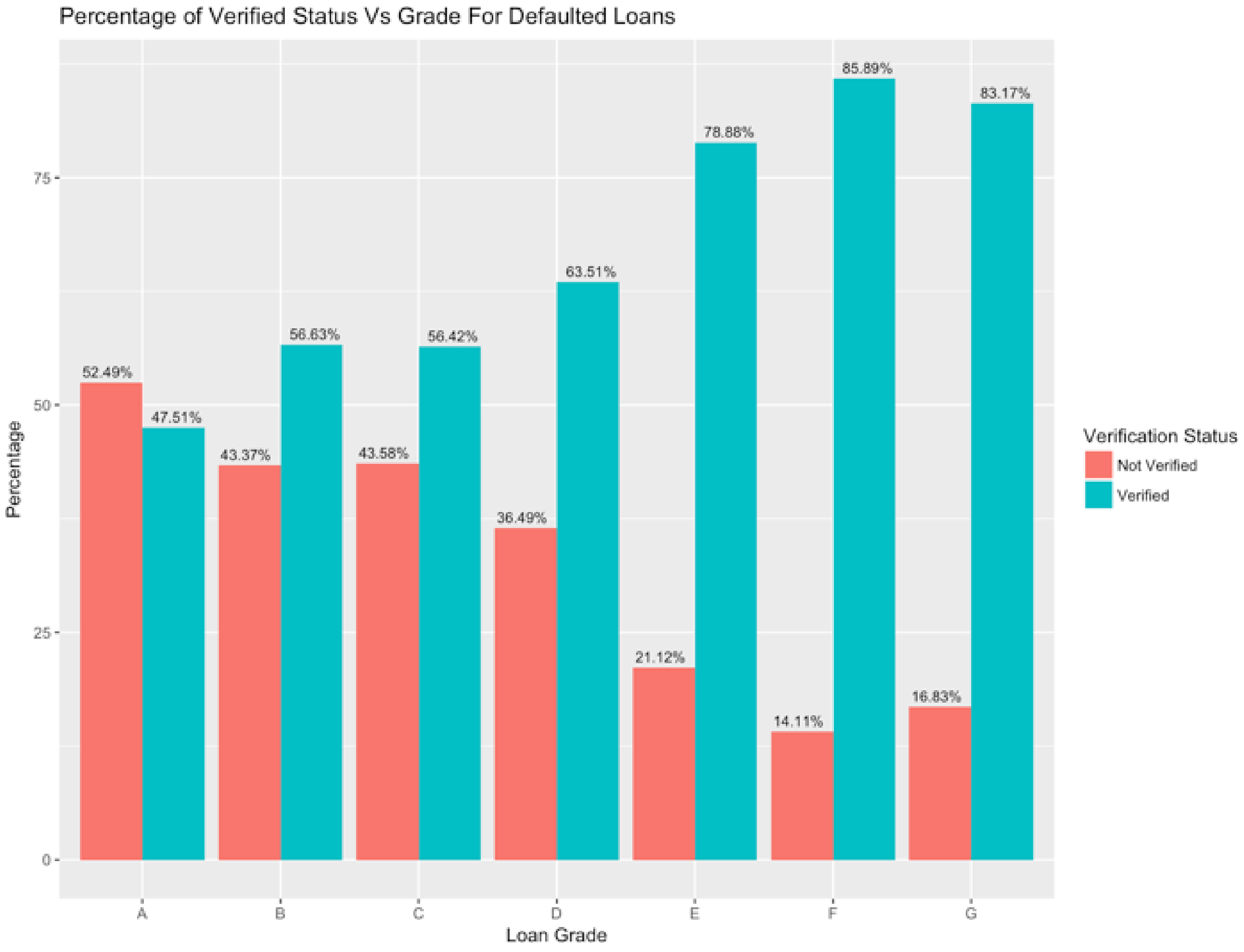
Observations –

- For the given dataset we see that there are a 62% of charged off loans post verification

Observations –

- For the given dataset we see that applicants whose income are verified default more across all loan products

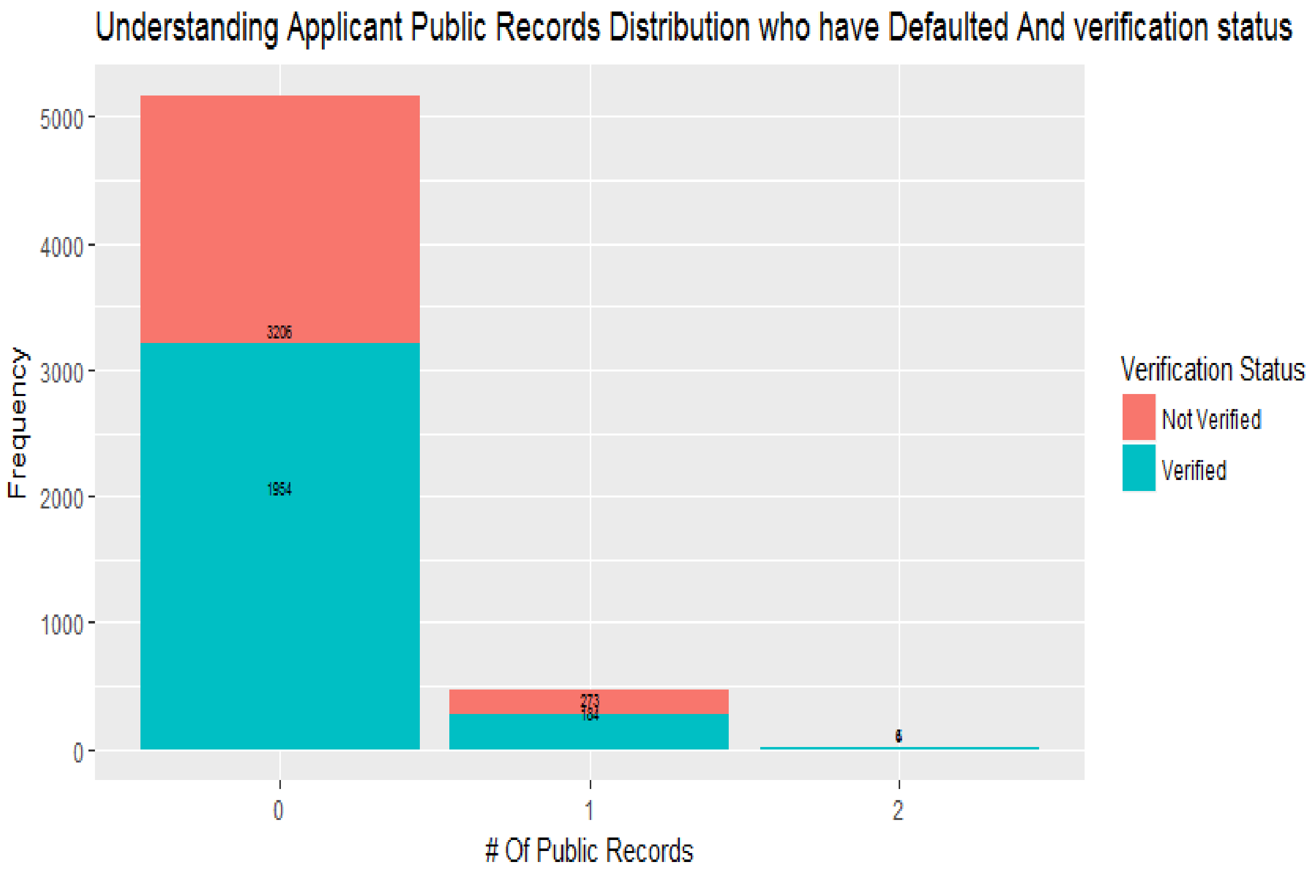
Distribution of charged off loans by grade and verification status



Observations –

- There is a higher percentage of Grade A, B, C loans which have not been verified. For such loans increasing the percentage of verification can reduce the risky applicants hence reducing the loan defaults.
- For the Grade E,F,G loans which comes with higher rate of interest (decide by other factors like credit fico score, DTI etc) the percentage of verified loans being defaulted is high. While assessing the loan applicants for such loan grades even if the status is verified it is better to reduce the loan amount funder or better reject the application.

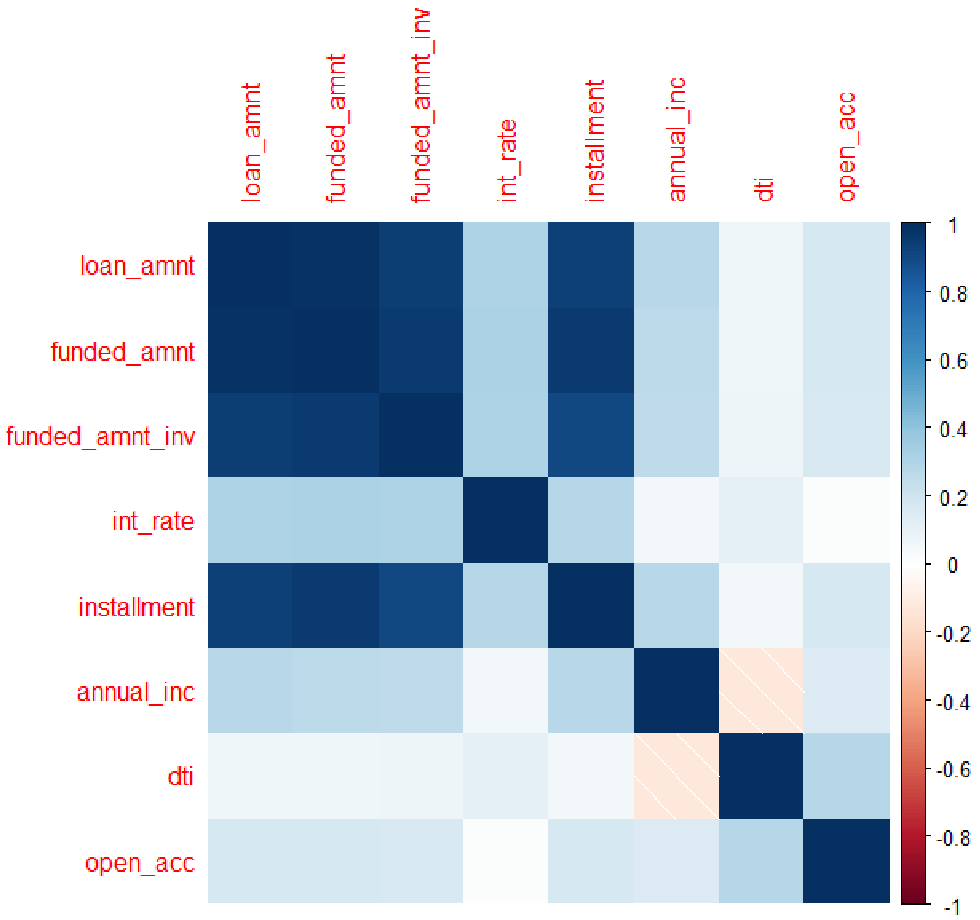
Distribution of charged off loans by grade and verification status



Observations –

- Even though the applicants who have zero derogatory remarks have defaulted however good set of them do not have income verified which makes them risky applicants and likely to default

Correlation Matrix for numeric variables



	loan_amnt	funded_amnt	funded_amnt_inv	int_rate	installment	annual_inc	dti	open_acc
loan_amnt	1	0.98	0.94	0.31	0.93	0.27	0.07	0.18
funded_amnt	0.98	1	0.96	0.31	0.96	0.27	0.07	0.18
funded_amnt_inv	0.94	0.96	1	0.31	0.91	0.25	0.07	0.16
int_rate	0.31	0.31	0.31	1	0.28	0.05	0.11	0.01
installment	0.93	0.96	0.91	0.28	1	0.27	0.05	0.17
annual_inc	0.27	0.27	0.25	0.05	0.27	1	-0.12	0.16
dti	0.07	0.07	0.07	0.11	0.05	-0.12	1	0.29
open_acc	0.18	0.18	0.16	0.01	0.17	0.16	0.29	1

Observations –

- There are high positive correlation between loan_amnt and funded_amnt, loan_amnt and funded_amnt_inv, loan_amnt and installment, funded_amnt and installment. There is slight negative correlation between dti and annual income which is obvious