Gramener Case Study

Business Objective and Strategy

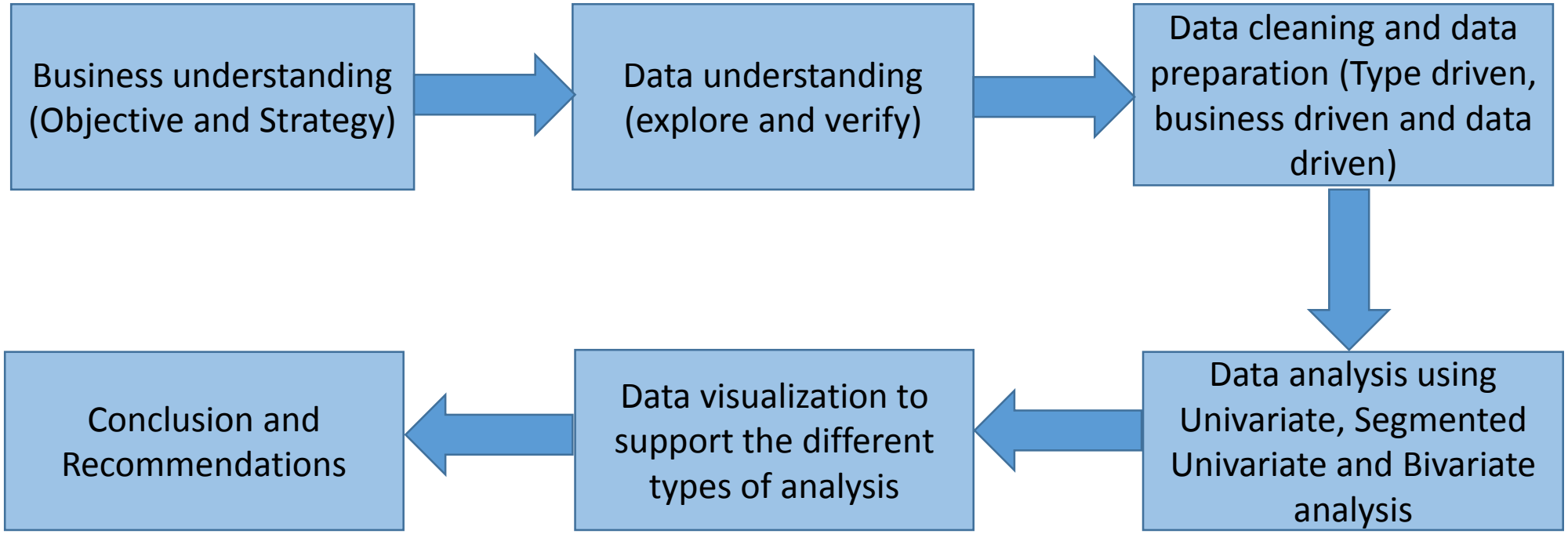
Business Objective

- The aim is to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.
- In other words, the company wants to understand the **driving factors (or driver variables)** behind loan default, i.e. the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.

Strategy

- 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
- How **consumer attributes** and **loan attributes** influence the tendency of default.

Problem Solving Approach



** These steps are detailed out and supported using graphs and plots in subsequent slides

Data Understanding (Explore and Verify)

Data

- Dataset gives data about the loans issued during the period 2007 to 2011
- There are 39717 observations of 111 Variables
- There are 2208180 N/A Values and 4408587 empty values
- There are no duplicate/repeated loan id or loan member id
- There are several variables which have only NA as values
- There are several variables which have only empty as values
- There are several variables which have only one Unique value

Metadata – information about few important variables

dti : A ratio calculated using the borrower's total monthly debt payments on the total debt obligations, excluding mortgage and the requested LC loan, divided by the borrower's self-reported monthly income.
inq_last_6mths: The number of inquiries in past 6 months (excluding auto and mortgage inquiries)
loan_status: Current status of the loan [current, charged off and fully paid]
Grade: A grade that is assigned to every loan. This grade is based on the credit history of the borrower
Verification_status: Indicates if income was verified by LC, not verified, or if the income source was verified

Data cleaning and preparation

- Remove variables where
  - All values are NA
  - All values are empty
  - Single unique value
- Remove variables which are not significant for analysis
  - url
- Remove the column which has 97% of N/A
  - next_pymnt_d

Data Preparation

- Some of the variables need treatment
  - int_rate variable has % associated with it. Remove the % symbol so that correlation with other variables could be found – bivariate analysis on continuous variables
  - Convert into factor variables for univariate and bivariate analysis
    - "term","grade","sub_grade","emp_length","home_ownership", "verification_status", "loan_status","purpose","zip_code"
- Outlier treatment for annual_inc variable – There are some values which are way outside the normal range
- Convert date variables into R accepted date formats
    - In some scenarios, we have to manually add “01” as day of the date to the fields
    - Split some date fields to year and month for analysis by year.
      - Issue_d
- Convert the numeric fields into two decimal places for consistency

Derived Metrics

Data driven metrics

- Some variables can be split and made into new categorical variables. These variables will help in plotting the data and getting insights
  - funded_amnt(0-35000) is split into 7 categories with a range of 5000 each
  - loan_amnt(0-35000) is split into 7 categories with a range of 5000 each
  - annual_inc(25000-150000) is split into 6 categories with a range of 25000 each
  - dti(0-30) is split into 6 categories with a range of 5 each
- A variable is created which will store the ratio of funded amount against the loan applied for

Business driven metrics

- Employment length is in years. Possible values are between 0 and 10 where 0 means less than one year and 10 means ten or more years.
  - Change the emp_length variable into accommodate this business rule

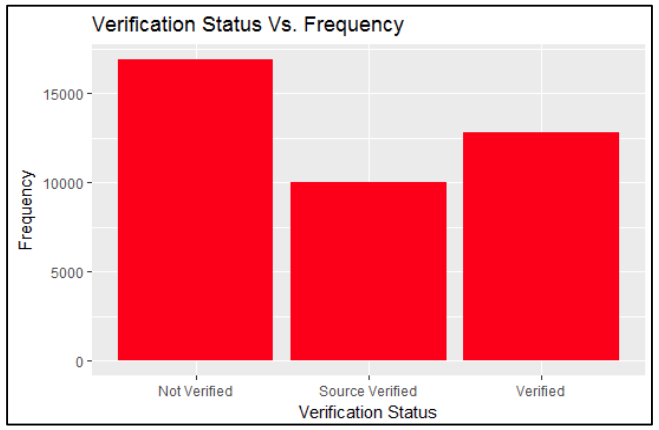
Type driven metrics

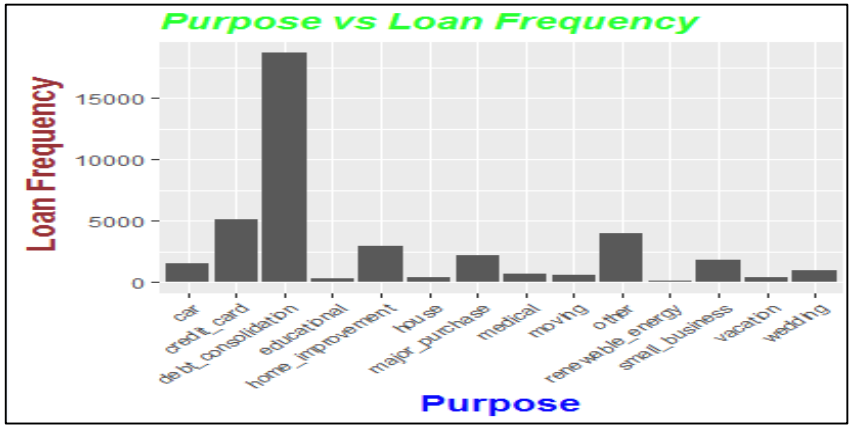
- A type driven metrics is derived in such a way that, if the purpose loan description is given in less than 50 characters, we assume that less details are provided or else we assume that good details are provided. This may help us in analyzing the charged off candidate

Univariate Analysis

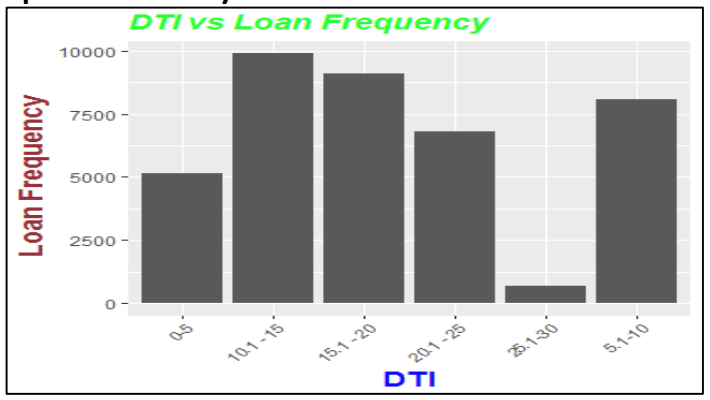
Univariate analysis on categorical variables

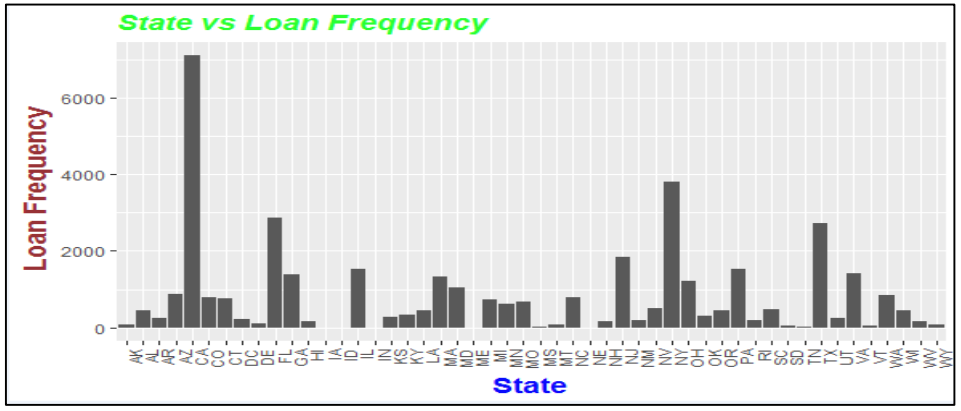
- Plot a graph to analyze the use case of verification status - It seems that “Not verified” cases are more in numbers. We will analyze more to find out what other parameters are affecting this
- Plot a graph to analyze the use case for purpose of loan – Looks like most Loans are for purpose - debt consolidation





- Plot a graph to analyze the use case for DTI - It seems most Loans are from the DTI between 10-20
- Plot a graph to analyze the use case for state address – Most borrowers are from CA state

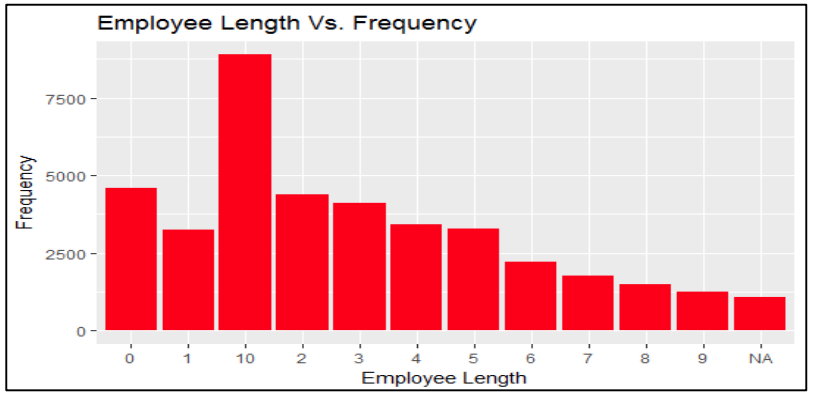


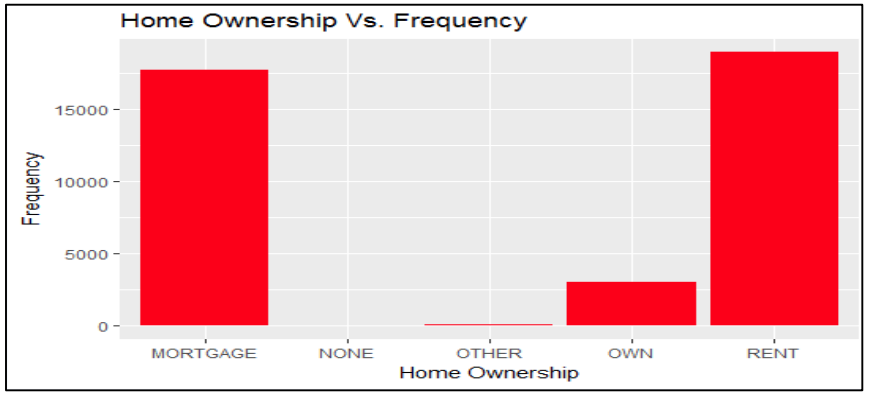


Univariate Analysis

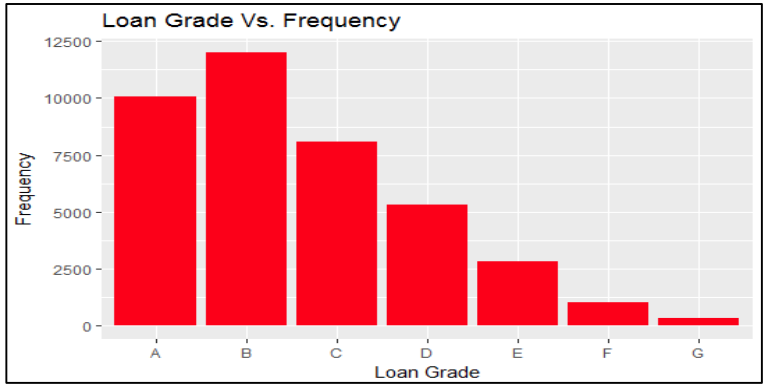
Univariate analysis on categorical variables – contd.

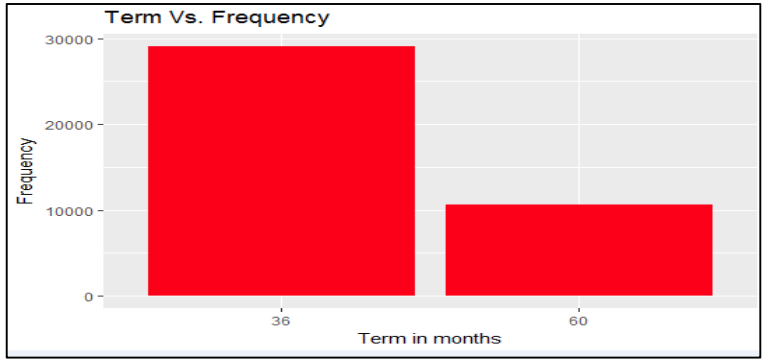
- Plot a graph to analyze the use case of Employee experience – It seems that people with 10 or more years of experience requested for loan heavily
- Plot a graph to analyze the use case for home ownership – People who stay in rented house requested for loan more than any other





- Plot a graph to analyze the use case for Loan grade- Grade frequency is in order of B,A,C,D,E,F,G
- Plot a graph to analyze the use case loan term- More people opted for 36 month term

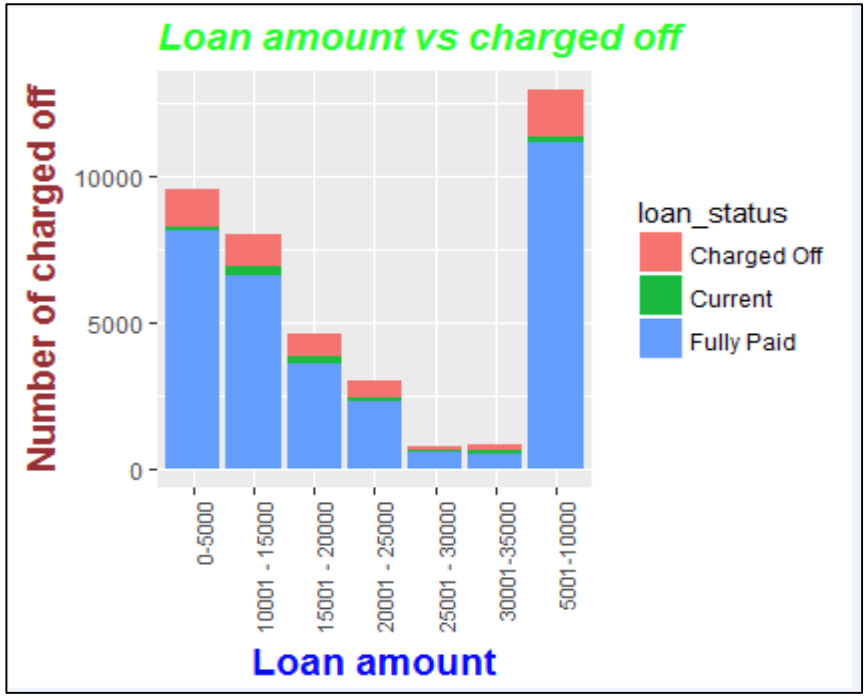


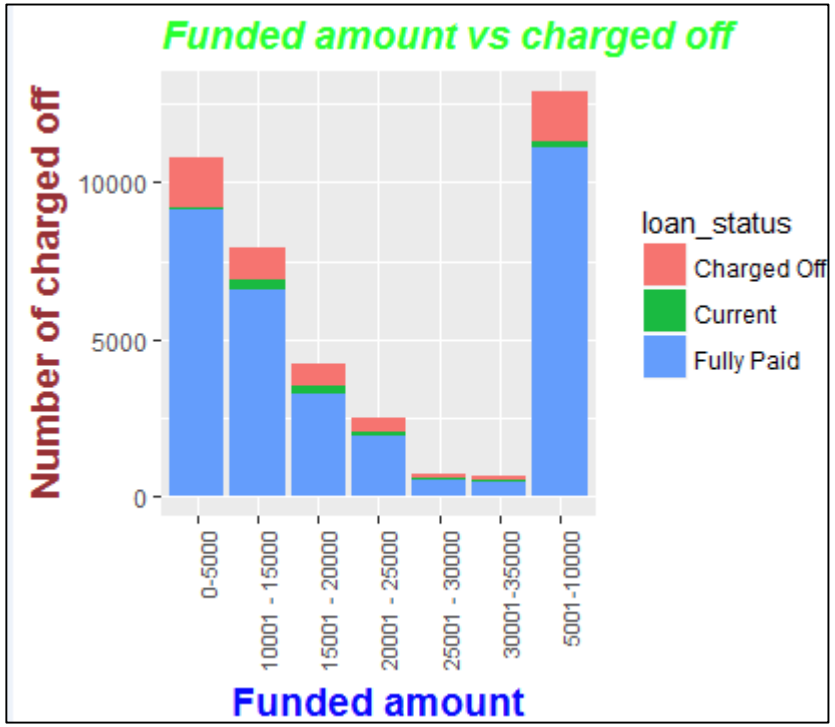


Univariate Analysis

Univariate analysis on categorical variables – some more insights

- It seems most charged off loans lie between the applied loan amount between 0-10000
- It seems most charged off loans lie between the funded amount 0-10000

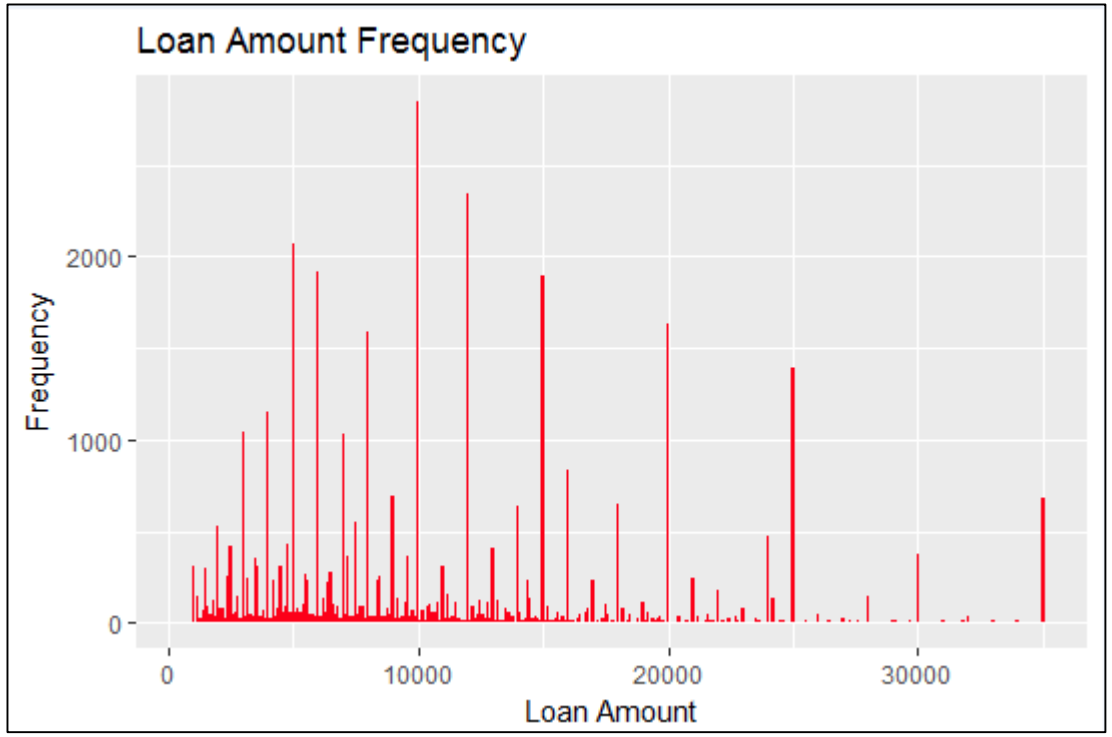


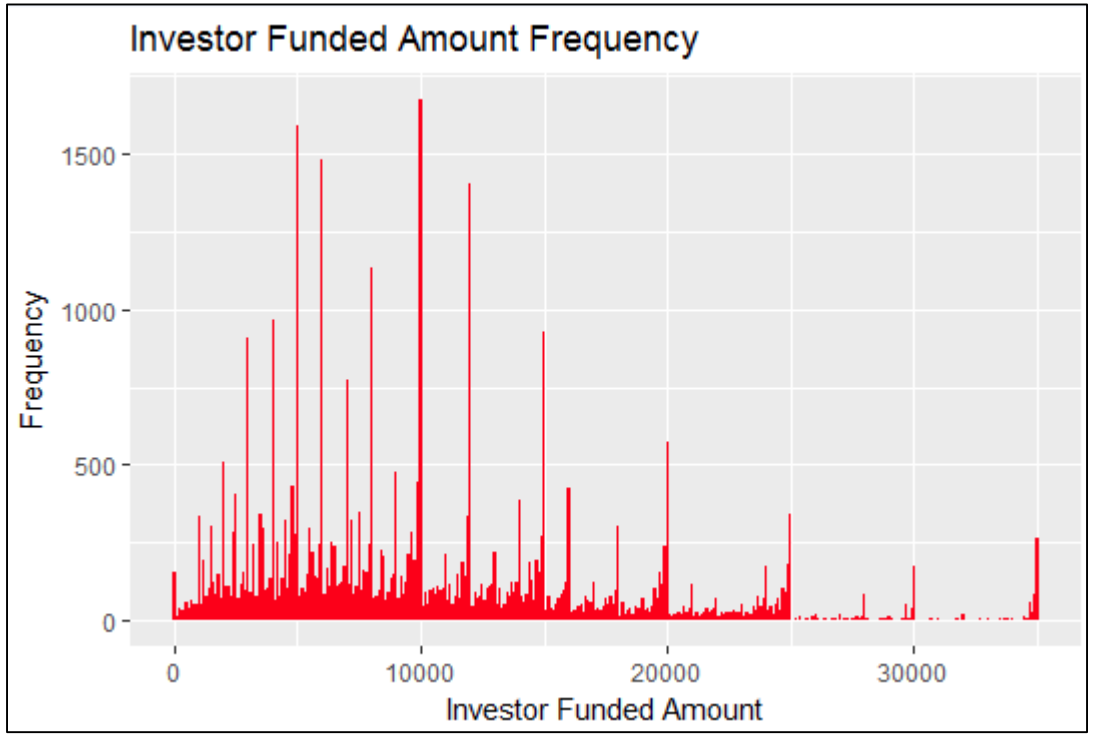


Univariate Analysis

Univariate analysis on continuous variables

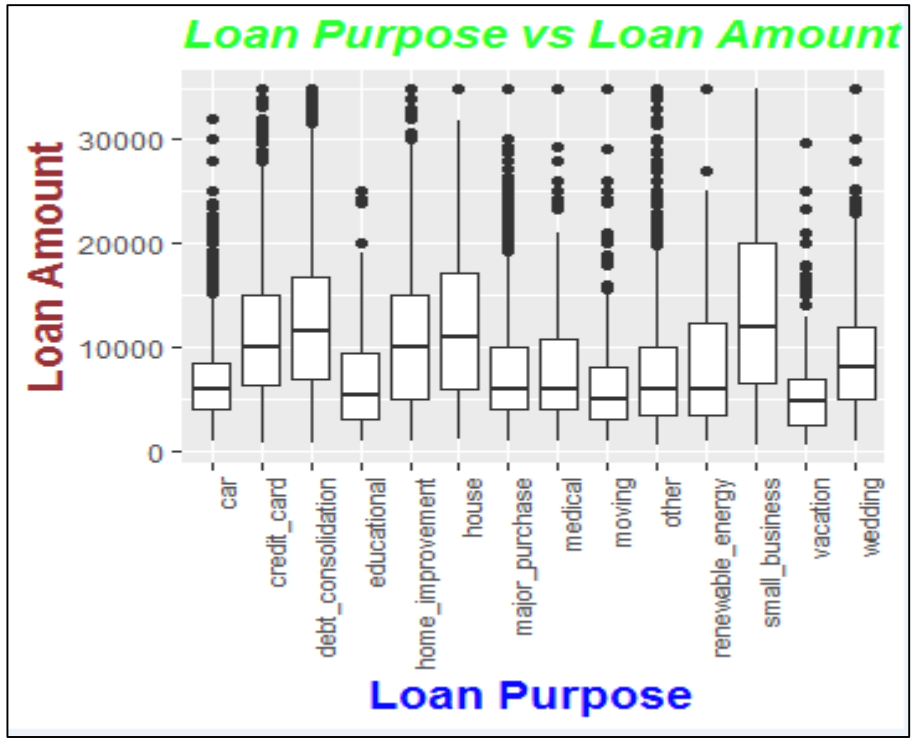
- Loan_amnt Mean is 11219.44 and Median is 10000 which are same to some extent but
- funded_amnt Mean is 10397.45 and Median is 8975 which are significantly different
- It seems that the amount that is requested by the borrower is almost fulfilled by the investors. This can be clearly seen from the two graphs

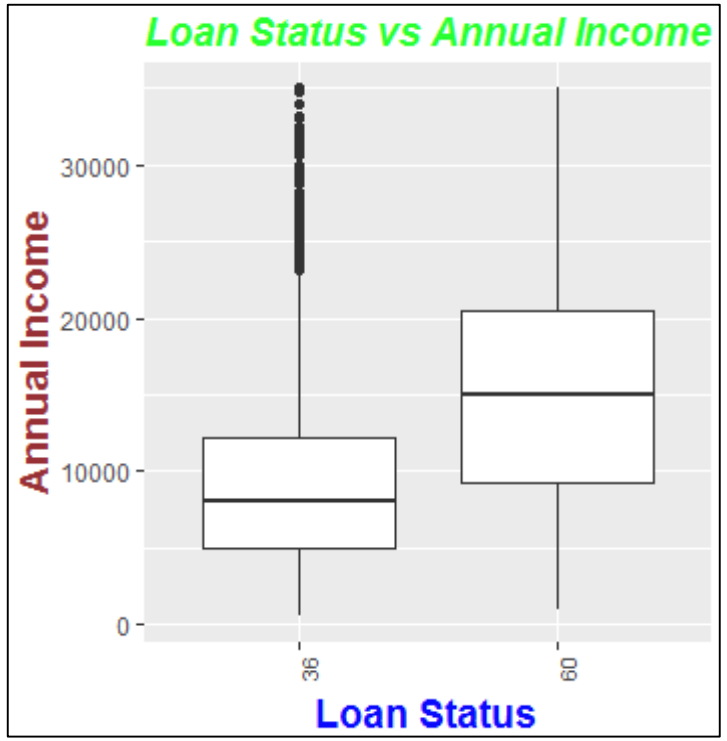


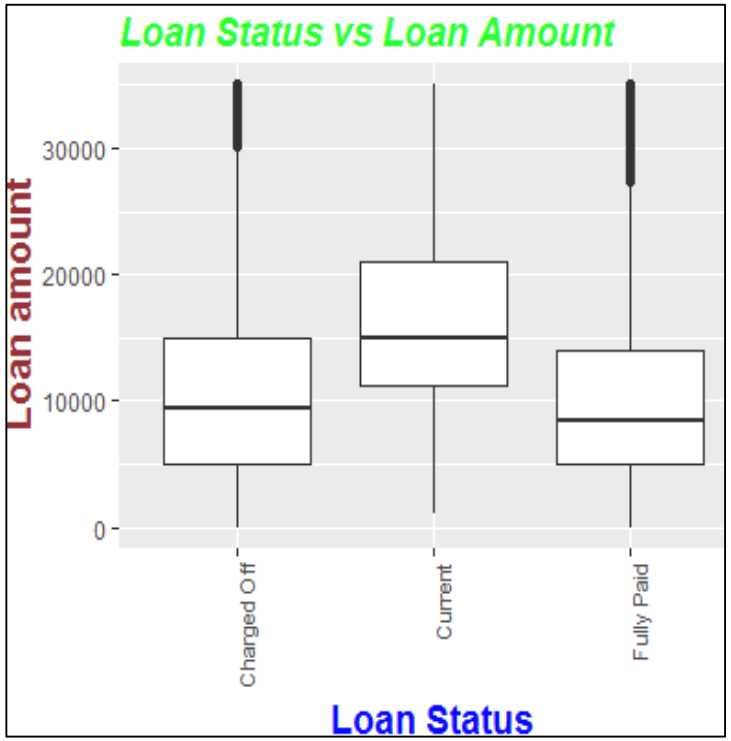


Segmented Univariate Analysis

- Small Business and Debt consolidation have the highest loan amounts across categories
- Surprisingly the 60 month term has a mean value much higher than that of the 36 month term loan
- Out of all the charged off loans, 50% of loan amount is less than or approx. equal to value of 10,000

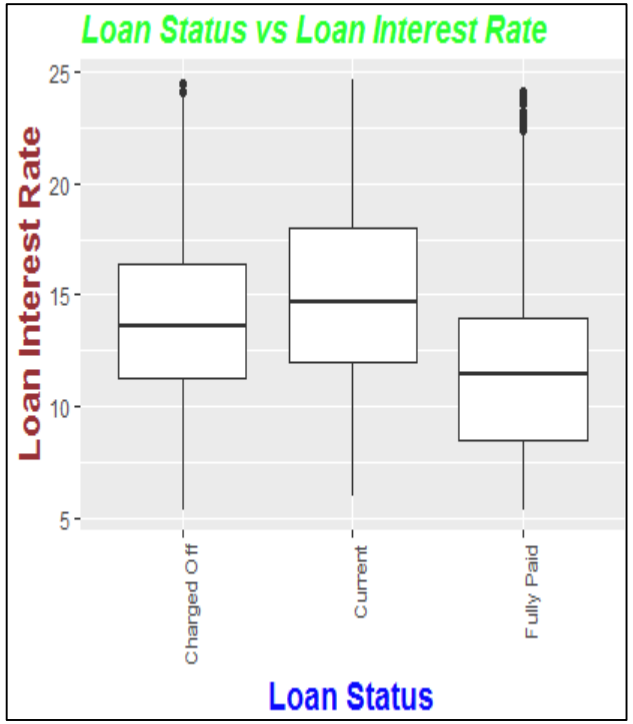


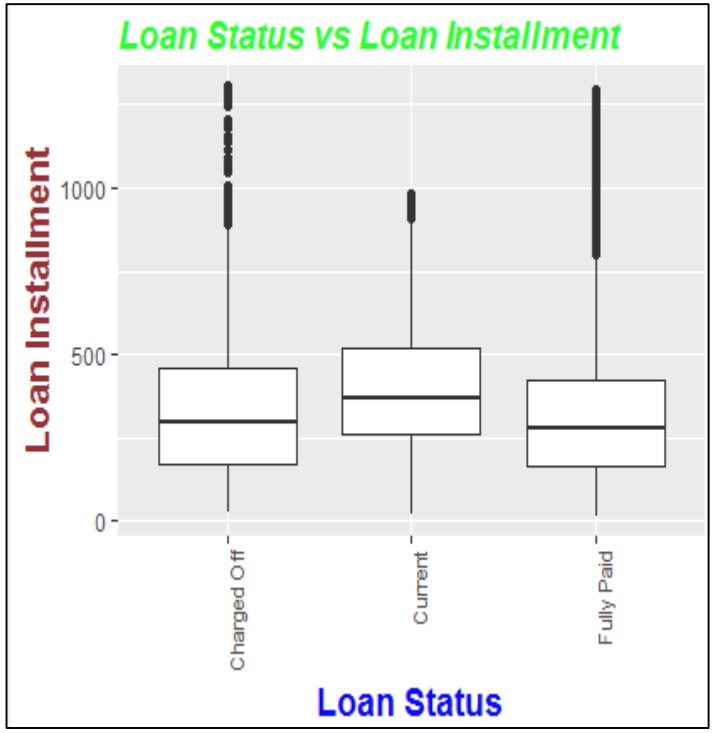


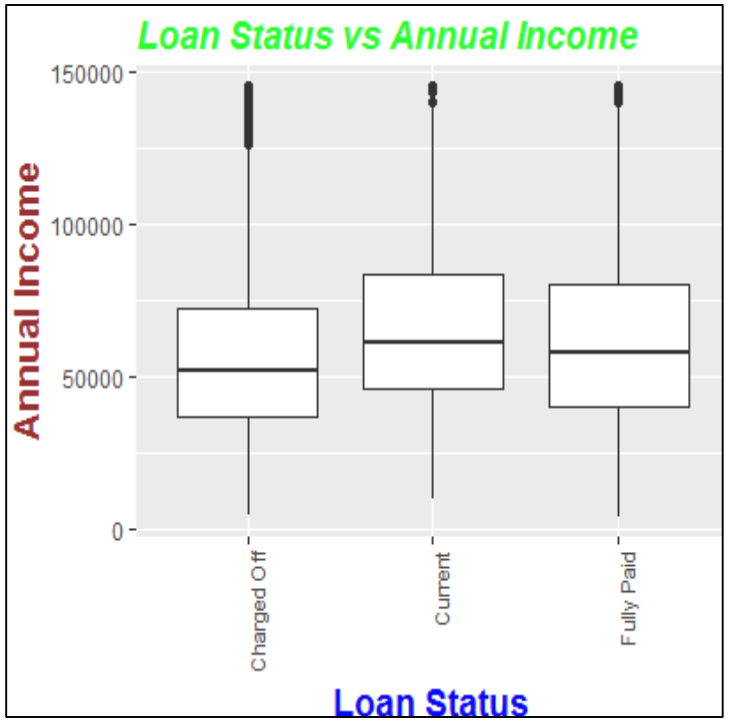


Segmented Univariate Analysis

- Out of all the charged off loans, 50% of loan interest rate is less than or approx. equal to 13.75
- Out of all the charged off loans, 50% of loan installment is less than or approx. equal to 300
- Out of all the charged off loans, 50% of annual income is approx. equal to 50000



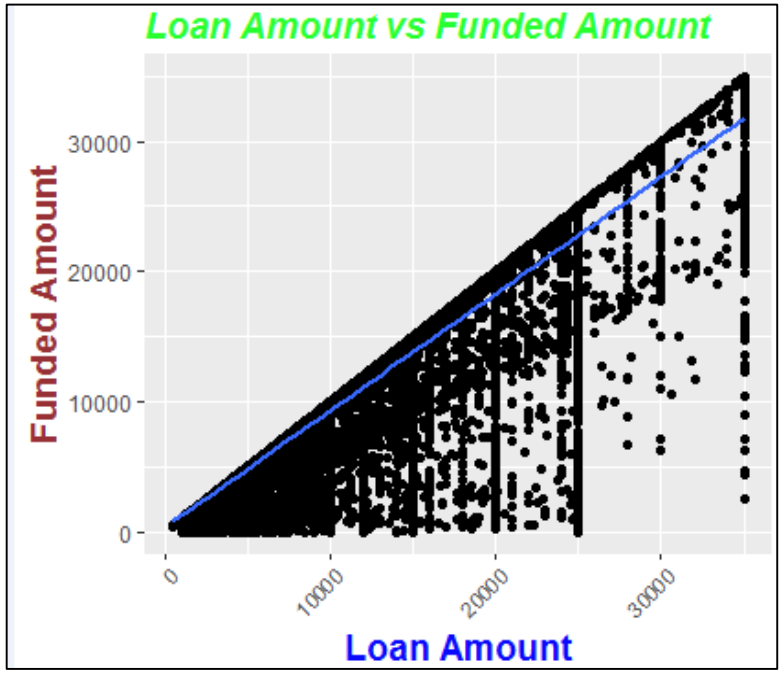


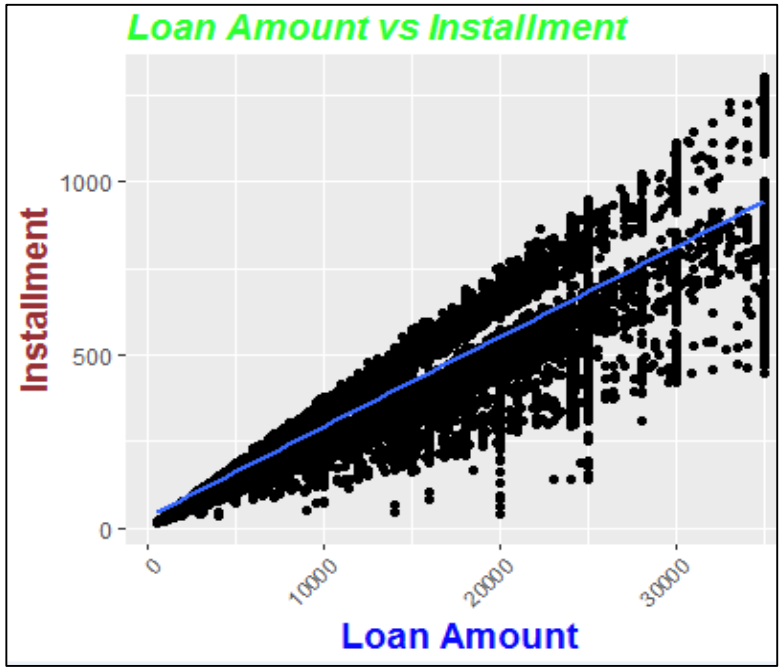


Bivariate Analysis

Bivariate analysis on continuous variables

- Correlation between loan amount applied and funded is .94 - This implies that the loan amount applied for is highly correlated with amount being issued.
- Correlation between loan amount approved and installment is .93 - This implies that the loan amount applied for is highly correlated with installment amount
- Correlation between interest rate and Debt to income is 0.1 It seems that change in debt to income value has no or less effect in interest rate.
- Correlation between annual income and Debt to income is -0.07 It seems that change in debt to income value has not effect in annual income



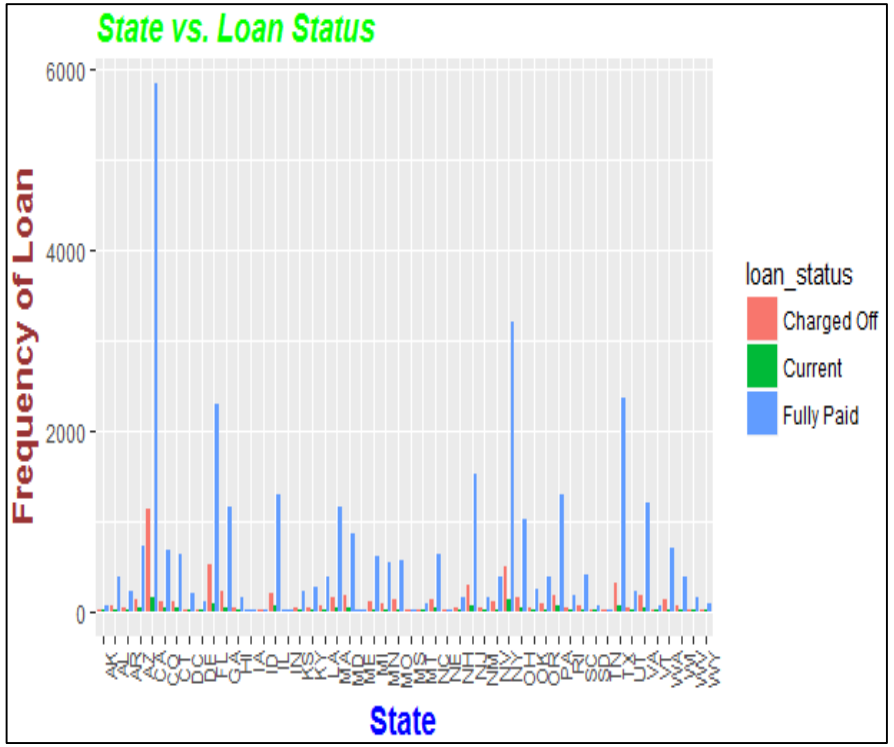


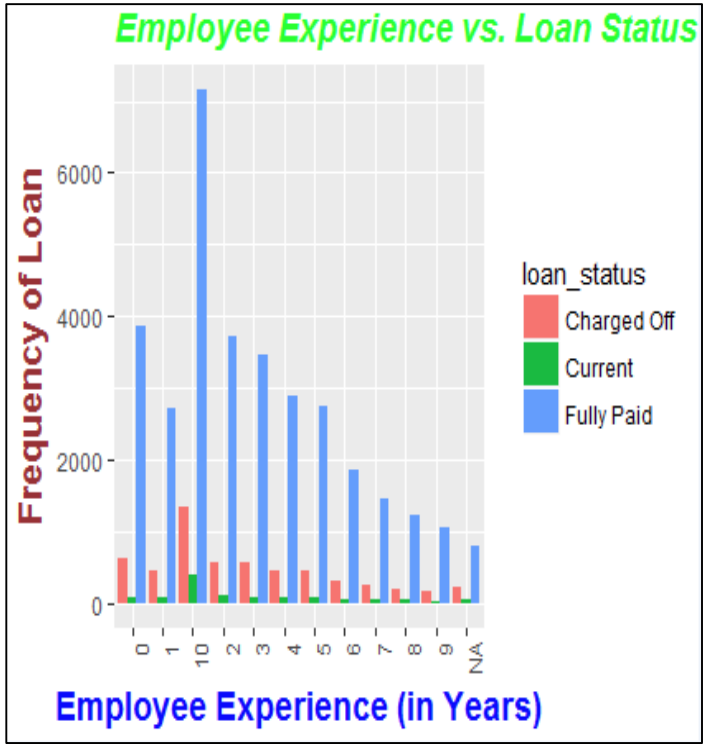
Bivariate Analysis

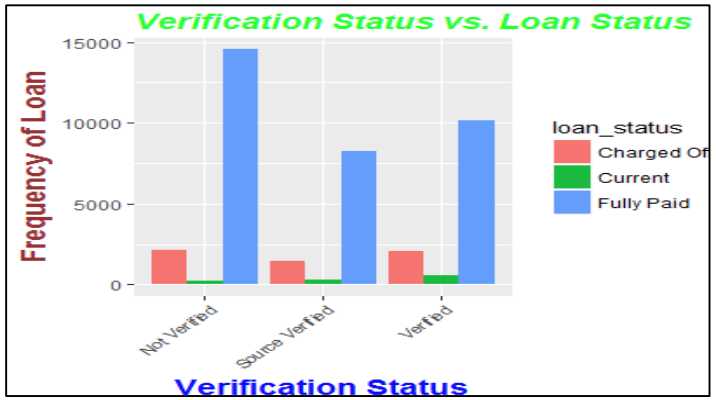
Bivariate analysis on categorical variables:

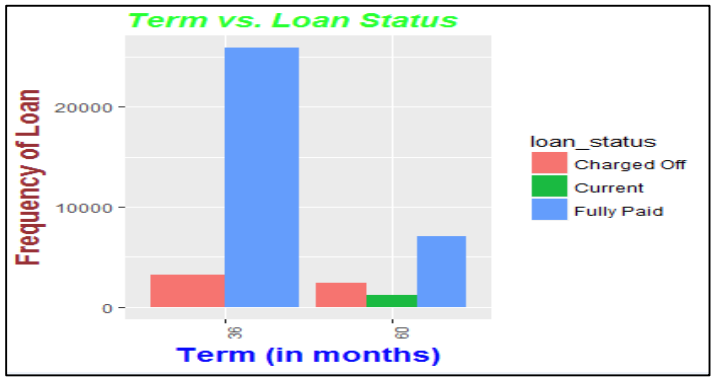
Let us plot some bar graph for loan status against different variables to find the possible defaulter parameters

- Defaulters vs. State - People staying in the state of CA are mostly defaulter
- Defaulters vs. Experience - People with experience of 10 or more years are mostly defaulter
- Defaulters vs. Verification status - People for whom verification is not done are mostly defaulter
- Defaulters vs. Loan term - People who opted or provided with 36 month loan term are mostly defaulter





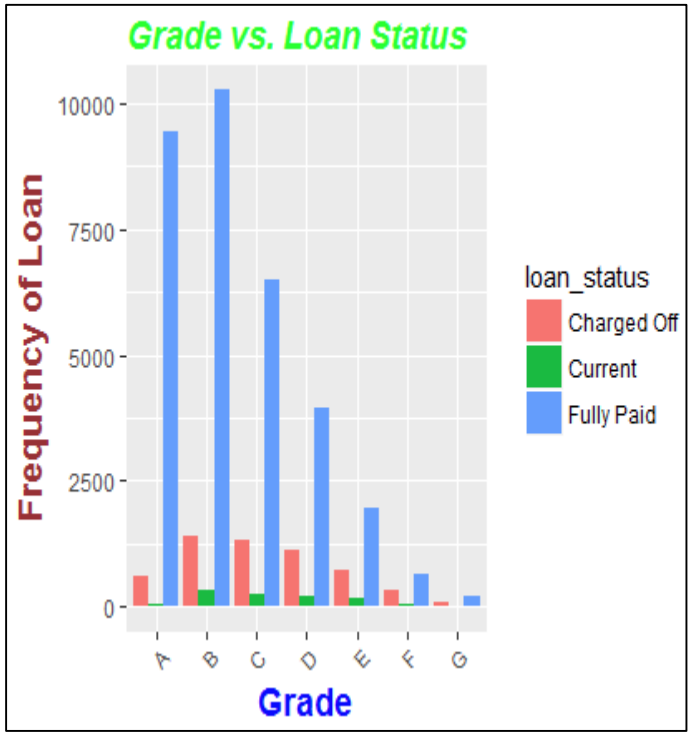


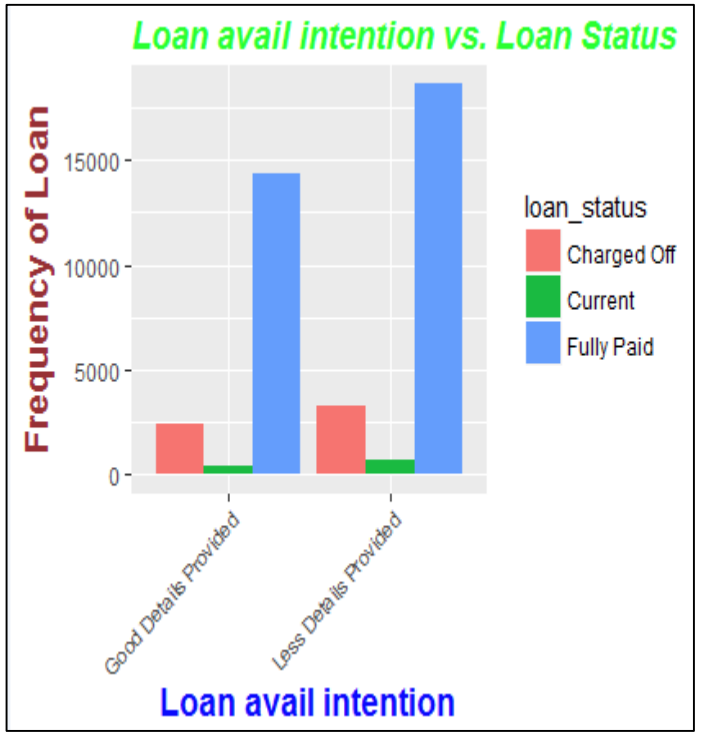


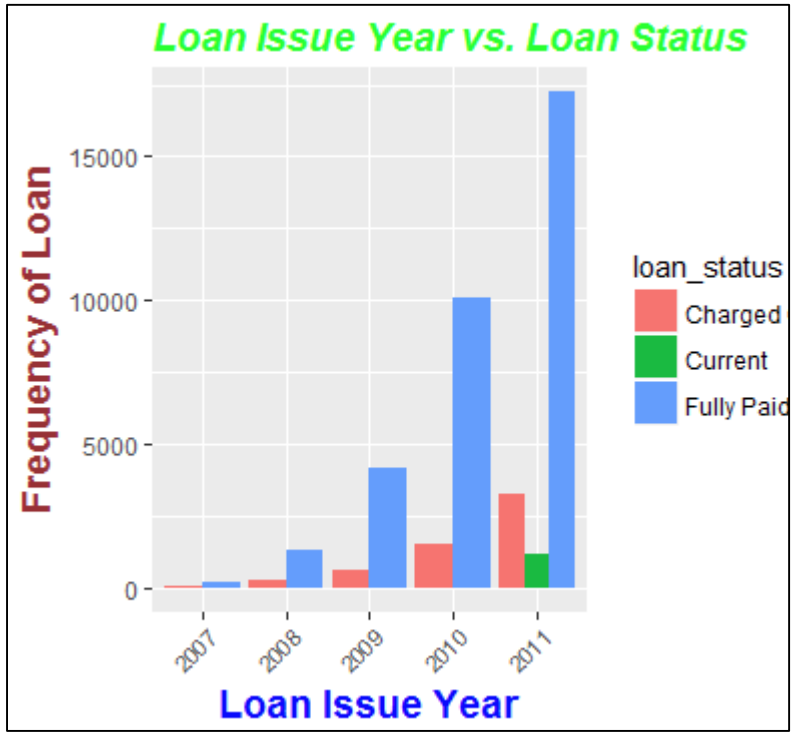
Bivariate Analysis

Bivariate analysis on categorical variables: contd.

- People who have been provided with loan under grade B & C are mostly defaulter
- People who have not provided enough details for the intention of loan are mostly defaulter. This is a type driven metrics variable
- Additional Insight : Mostly the number of defaulters are increasing year by year







Insights

Insight 1:

Few important driver variables (i.e. variables which are strong indicators of defaulter)

- Loan Verification Status = Non Verified
- Loan Term = 36 month
- Borrower's experience >=10
- Loan Grade = B or C
- Intention of Loan = Less description about the loan purpose
- State = CA

Insight 2:

- It is found that Percentage of charged off loan for non verified cases is 38% that is more than verified cases of 36%.
  - But when experience level of 10 or more years is considered, we determined that 30% of borrowers are defaulters even though they are verified compared to 20% of borrowers who are defaulters even when they are not verified
    - **It feels like non verified borrowers will mostly be the defaulters but it is also determined that verified customers are also defaulters if the experience level is more than 10**

Insight 3:

- It is found that Percentage of charged off loan for non verified cases is 38% that is more than verified cases of 36%. Also people who opted for 36 month term will mostly be defaulter
  - But when term 60 month is considered, we determined that 58% of borrowers are defaulters even though they are verified compared to 23% of borrowers who are defaulters when they are Not verified
    - **It feels like non verified borrowers will mostly be the defaulters but it is also determined that verified customers are also defaulters if the term is 60 month**

Insights – Contd.

Insight 4:

- It is found that 43% and 57% of defaulters are having a loan term of 60 and 36 month respectively.
  - Also people who opted for loan are mostly under Grade B and Grade A
    - **It seems defaulters are mostly from grade B and with 36 month term - 985 borrowers**
    - **It seems defaulters are mostly from grade E and with 60 month term - 539 borrowers**

Insight 5:

- It is found that borrowers with experience of 10 years are more prone to default than anyone else followed by borrowers with 0 years of experience
  - Also the percentage of defaulters for 10 years is 24% and 0 years is 11% but when home ownership is considered then defaulters with 10 years is 36% and defaulters with 0 years is 64%
    - **It is true that borrowers with experience level of 10 or more are risk to loan, however if they have a rented house then borrowers with experience level of 0 year are more risk to the loan than 10 years**

Recommendations

1. Below are the few important driver variables (i.e. variables which are strong indicators of defaulter). Before we can approve a loan, these factors should be consider

- A loan should be completely verified
- It is advisable to make the term as 60 month so that more time is given and installment can be less
- Be careful when approving loan for people with experience of 10 years or more. However if they have own house or rented house then scenario is different
- Loan given under grade B and C is most prone to risk. Make sure to check the credit history and Debt to Income ratio
- Make sure that borrower mentions the intention of loan. A good detailed purpose is always a good indicator
- More defaulters are from CA state, so be considerate and careful while giving loan to applicants belonging to CA state

Recommendations contd.

It is not advisable to directly deny the loan application based on the identified driver variables. Sometimes it may be a good idea to reduce the loan amount or increase the rate of interest to reduce the risk of losing business

2: Reducing the amount of loan

Any of your debts that reports to credit bureaus would count in your debt to income ratio, along with your new mortgage payment. Debts to consider would include student loans, credit cards, auto, and any other installment payments. Payment towards your rent would not be considered during the DTI ratio

- Business Rule assumption : In general any DTI less than or equal to 36% is considered as ideal. Any DTI above 36% is not ideal and will be a candidate for loan amount reduction. In our dataset, considering the DTI is between 0 and 30:
  - Any DTI less than or equal to 18% is ideal irrespective of the home ownership status, so suggestion is to approve the loan without any reduction in amount
  - Any DTI greater than 18% and less than 36% but with rented house is risk to loan as the installment might impact. Hence in this case, suggestion is to reduce the loan amount requested for
  - Any DTI greater than 18% and less than or equal to 36% but not with rented house is also ideal, so suggestion is to approve the loan without any reduction in amount
  - Any DTI greater than 36% is risk to the loan irrespective of home ownership status, so suggestion is to reduce the loan amount

Recommendations contd.

3: Increase the rate of interest

- If in the credit history of applicants, it is found that the person opting for the loan had several credit enquiry then this implies that this person might have several other loans and may be trying to reduce his/her current loan APR by opting new loan.
  - In this case, provide loan to these applicants but with a higher interest rate to compensate the loss , if any
- In our case, we can consider that if number of credit enquiry in last 6 month is less than or equal to 3, it is idle case otherwise provide loan with increase in rate of interest