GRAMENER CASE STUDY

SUBMISSION

Suman Ashwin (DDA1710279)

Gramener loan case study

Business Understanding:

- Lending club is an online market place to enable peer to peer lending of money between the borrowers and lenders. Borrowers are individuals who come to the LC to avail loans at a lower interest rate. Investors who lend on the LC get solid returns based on the interest rate they choose to lend.
- Like most other lending companies, lending loans to ‘risky’ applicants is the largest source of financial loss (called credit loss). The credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed. In other words, borrowers who default cause the largest amount of loss to the lenders. In this case, the customers labelled as 'charged-off' are the 'defaulters’

Business Objective:

- 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 utilise this knowledge for its portfolio and risk assessment.

Data categorization:

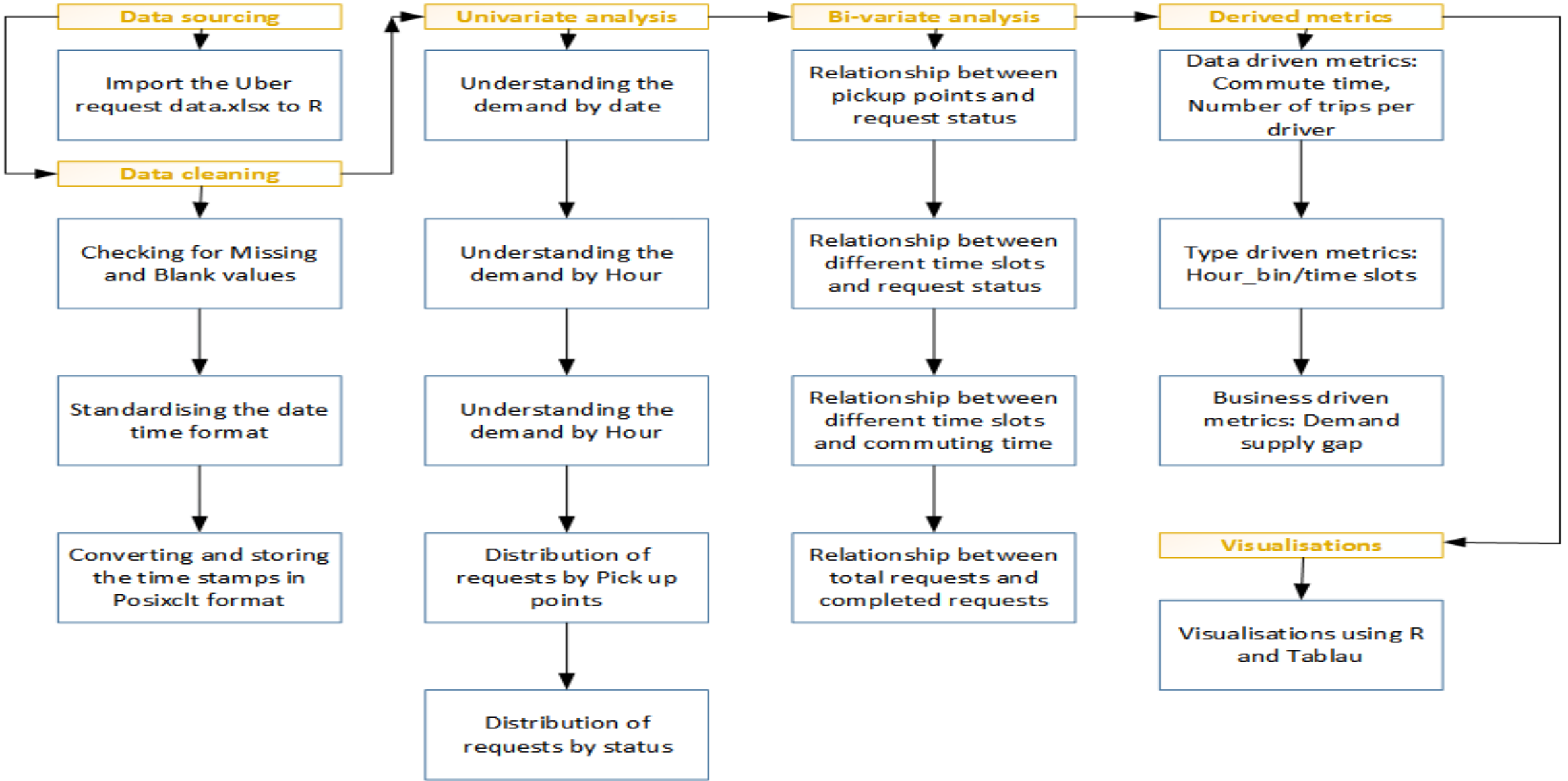
Loan accepted: If the company approves the loan, there are 3 possible scenarios described below:
Fully paid: Applicant has fully paid the loan (the principal and the interest rate)
Current: Applicant is in the process of paying the instalments, i.e. the tenure of the loan is not yet completed. These candidates are not labelled as 'defaulted'.
Charged-off: Applicant has not paid the instalments in due time for a long period of time, i.e. he/she has **defaulted** on the loan

Meta data available for the analysis

These are the list of variables we have got after the data cleaning procedure:

Loan Characteristics	Borrower Characteristics	Credit line Characteristics
Loan_amnt	Emp_length	Earliest_cr_line
Funded_amnt	Home_ownership	inq_last_6mths
Funded_amnt_inv	Annual_inc	open_acc
term	Verification_status	revol_bal
installment	purpose	revol_util
Int_rate	Addr_state	total_acc
Issue_d	dti	total_pymnt_inv
		total_rec_prncp
		total_rec_int
		last_pymnt_d
		last_pymnt_amnt
		last_credit_pull_d

Problem Solving Methodology



Approach for solving the case study

We are breaking down our analysis in following ways:

1. Identify the Loan, Borrower, Credit line Characteristics.

2. Bin the continuous variables into categorical variables for each of the 3 segments

3. . To establish the level of dependency with respect the loan status (Fully paid or Charged off), perform the proportion test and extract the chi square statistic

4. Compare the chi square statistic within each segment and pick the top 5 variables with the least chi square statistic, lower the chi square statistic number , higher is the dependency.

5 .For the chosen variables , perform the ggplot visualization to know the various attributes of the variable.

6.Establish a Pearson correlation matrix and corplot for all the numeric variables within a given segment

Chi-statistic numbers for different segments

Loan Characteristics

	chi_statistic
term.statistic	1159
loan_amnt.statistic	1335
funded_amnt.statistic	1497
grade.statistic	1569
sub_grade.statistic	1683
int_rate.statistic	2346
funded_amnt_inv.statistic	9370
installment.statistic	17926

Borrower Characteristics

	chi_statistic
homeown_risk.statistic	17.21
dti_cat.statistic	67.99
emp_length.statistic	71.62
erification_status.statistic	94.92
addr_state.statistic	156.72
annual_inc_cat.statistic	161.36
purpose.statistic	385.70

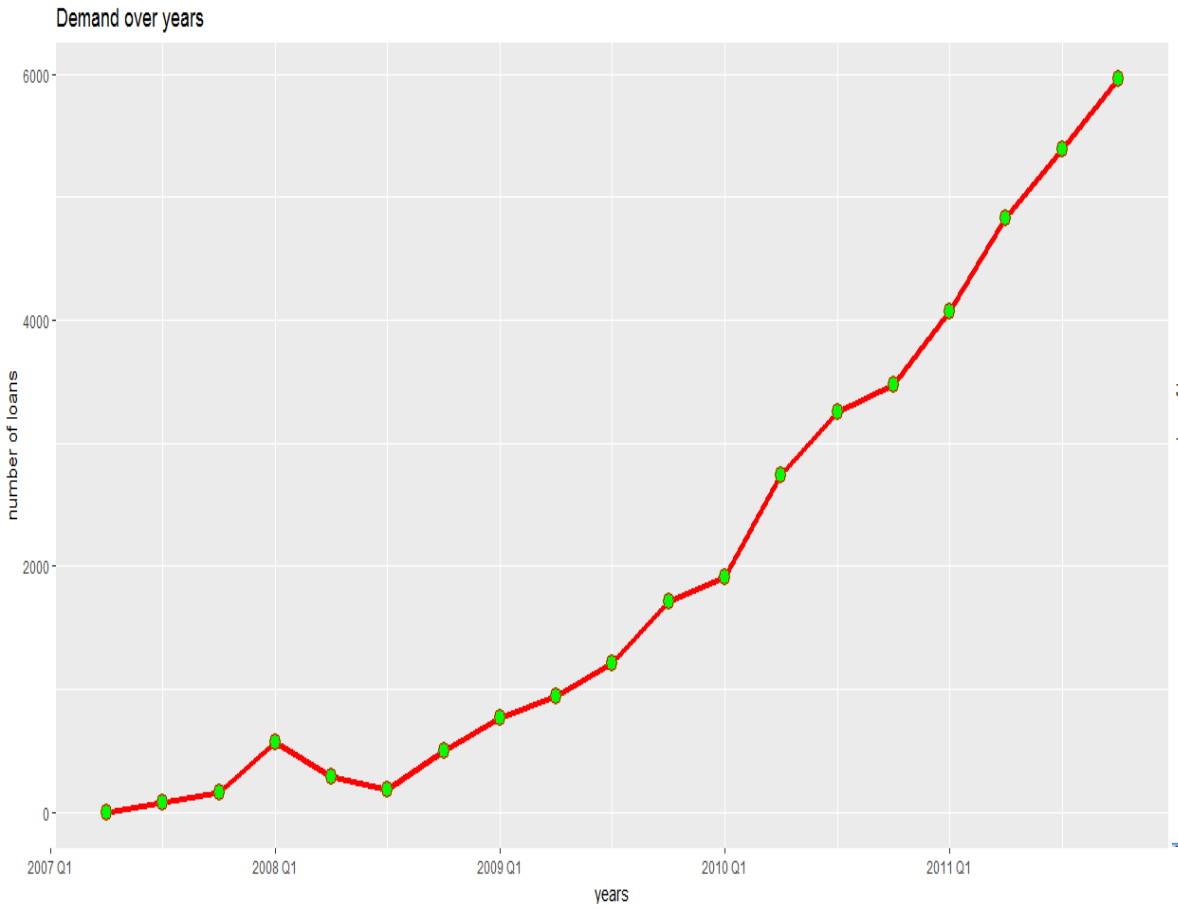
Credit line Characteristics

	chi_statistic
revol_bal_cat.statistic	10.15
total_rec_int_cat.statistic	26.95
total_acc_cat.statistic	33.45
open_acc_cat.statistic	35.37
inq_last_6mths_cat.statistic	133.10
revol_util_cat.statistic	243.52
last_pymnt_amnt_cat.statistic	927.58
total_pymnt_inv_cat.statistic	1521.07
total_rec_prncp_cat.statistic	2901.25

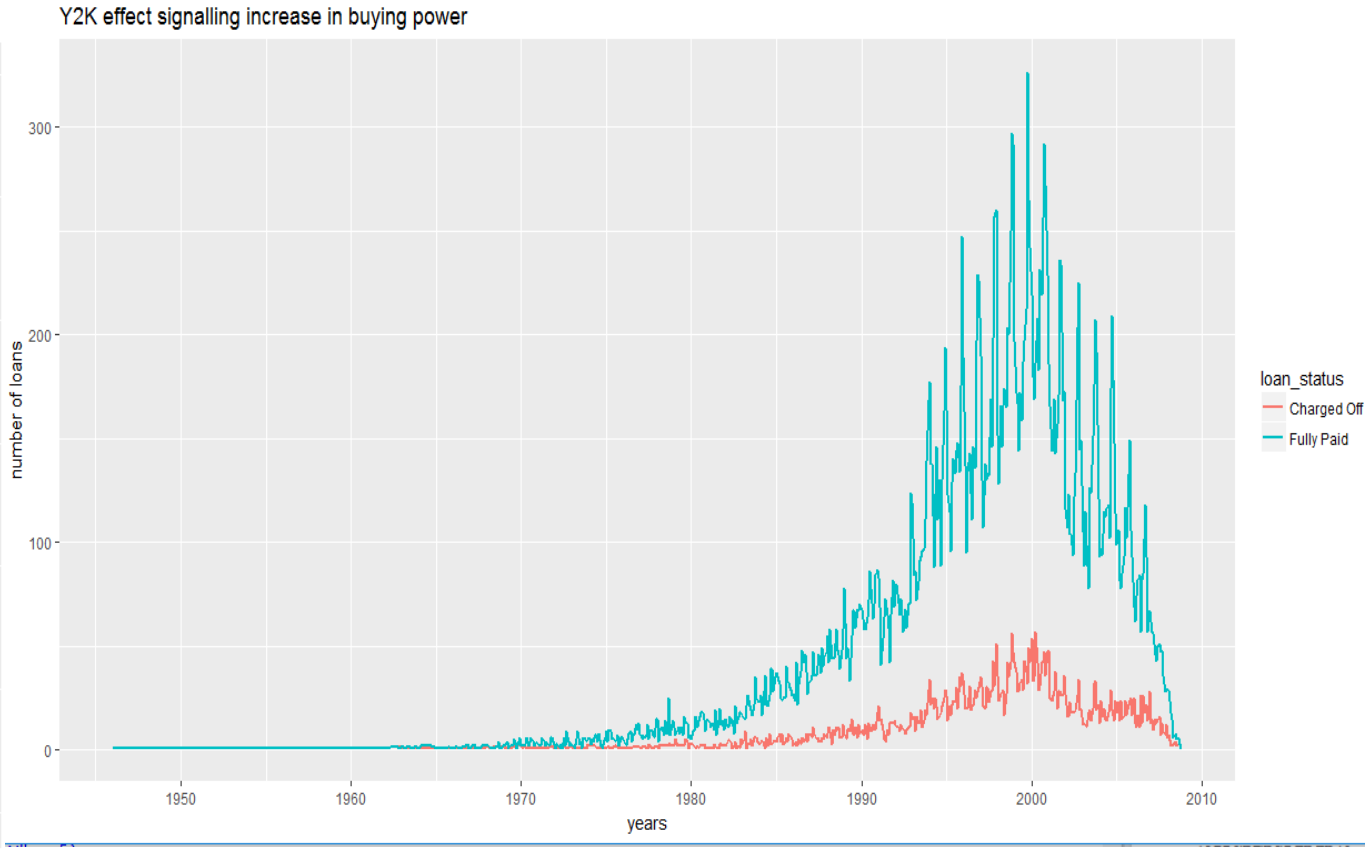
We will be considering the top 5 variables within each segment (highlighted in color) for further Univariate and Bivariate analysis

Generic analysis

Plot 1



Plot 2

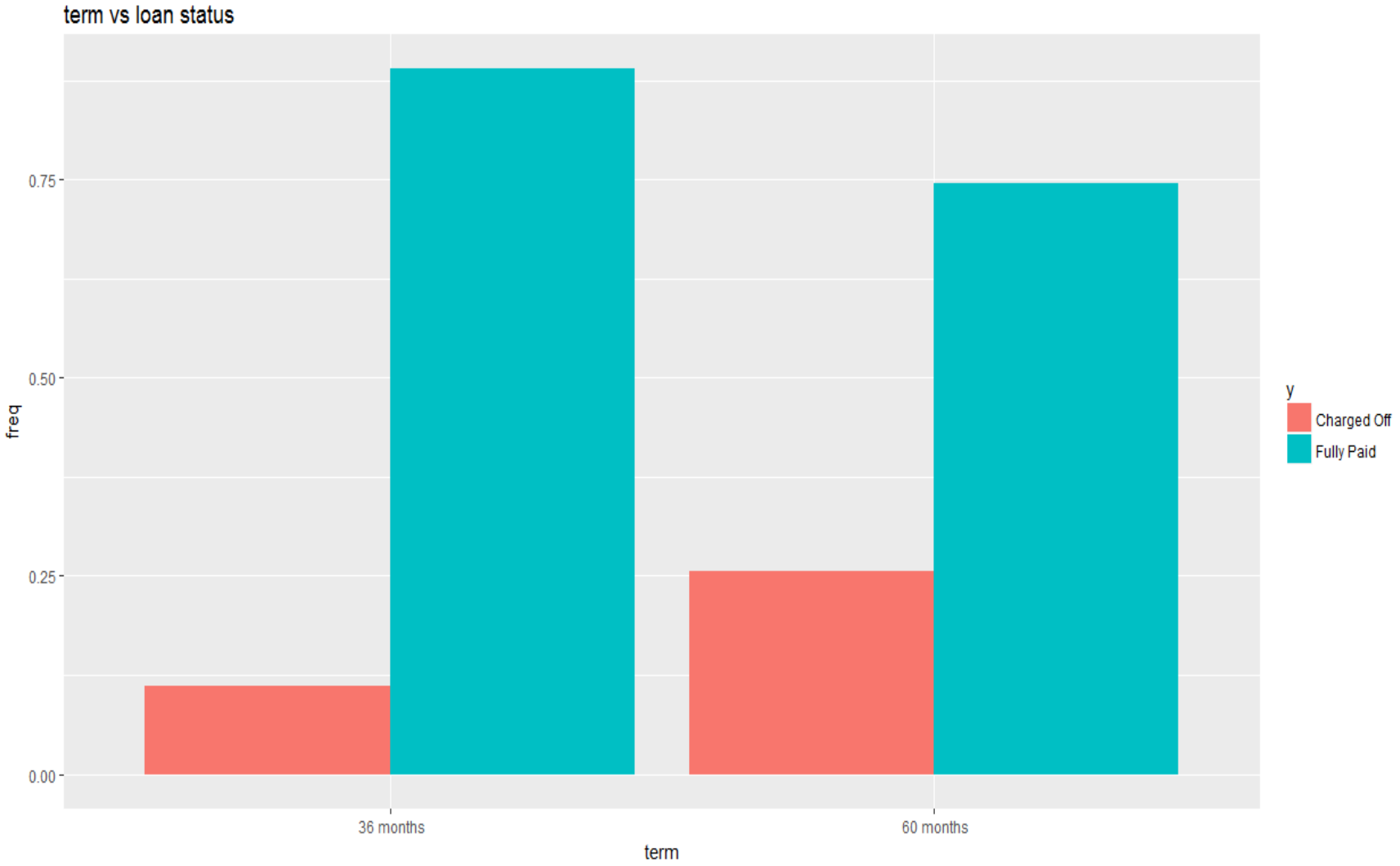


Findings

- The first plot Shows an exponential increase in the number of loans issued over by quarters by years
- In the second plot ,there is a spike in the number of credit lines opened in borrower account in the Year 2K, possibly coinciding with economic growth and the dotcom bubble

Analysis for the loan characteristics

1) term: The impact of tenure(36 months or 60 months on the loan status)



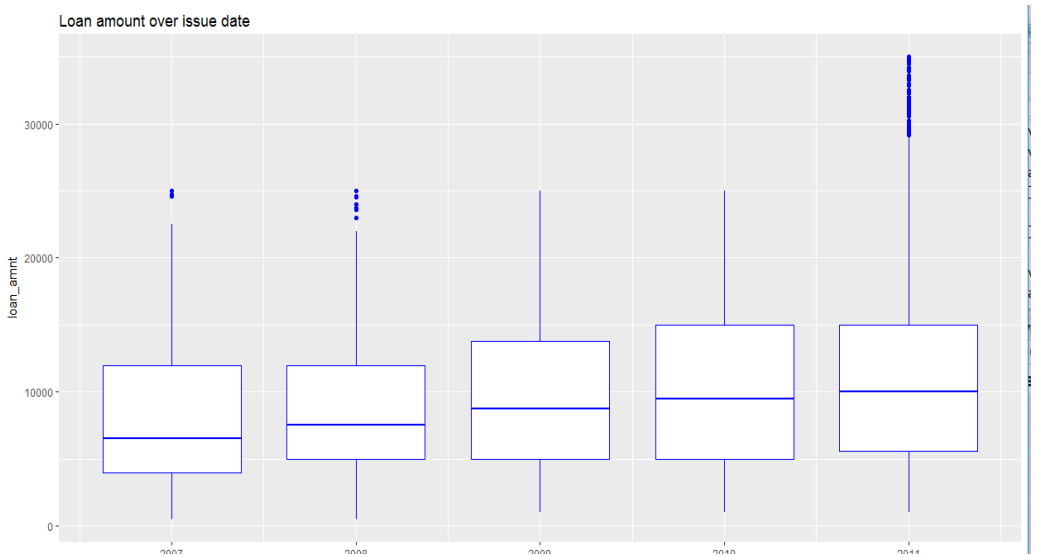
Findings

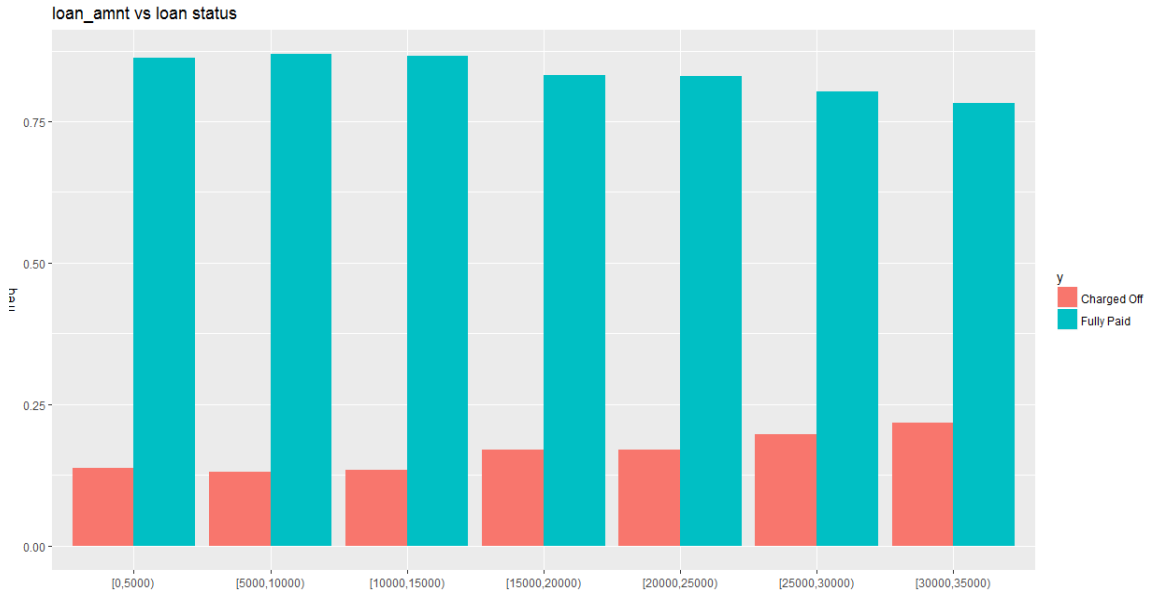
- #Shows the probability of a borrower defaulting is high(0.25) for 60 months tenure

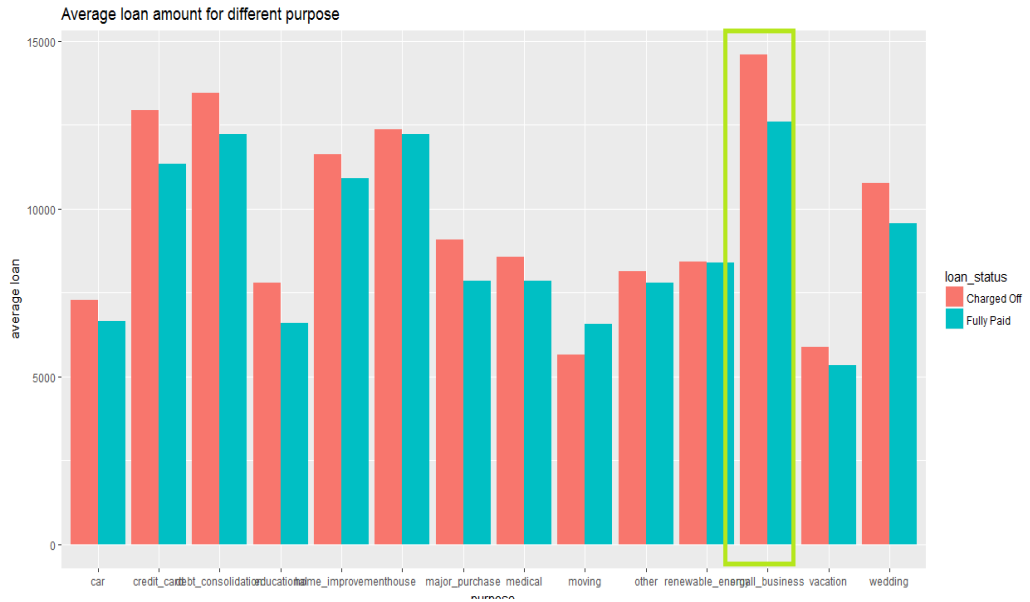
#	Charged Off	Fully Paid
# 36 months	0.11	0.89
# 60 months	0.25	0.75

Analysis for the loan characteristics(cont.)

2) loan_amnt :The impact of the loan_amnt on loan status





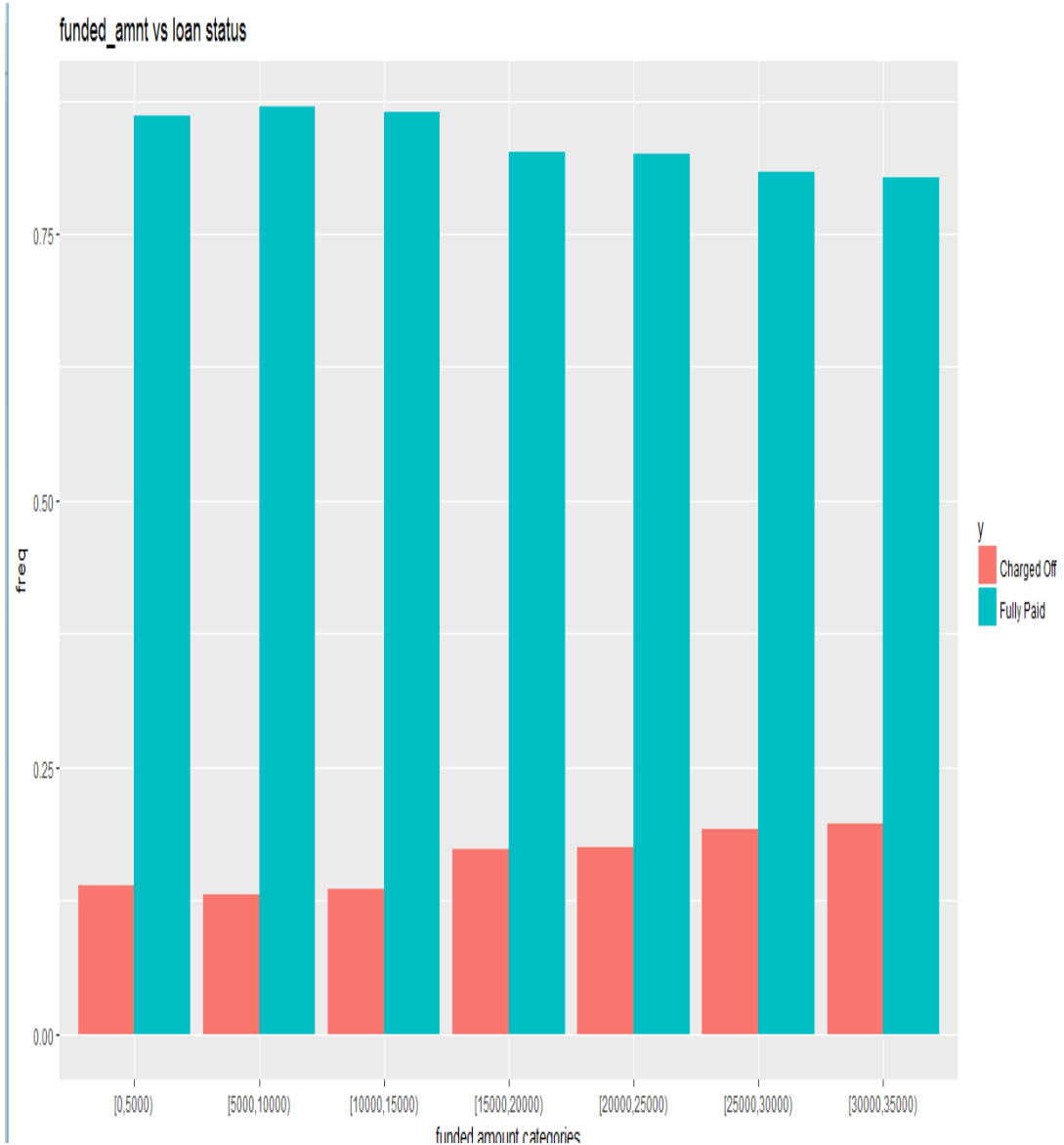


Findings

1. There is a marginal increase in the median loan amount seeked, over the years
2. Probability of borrower defaulting is relatively higher (0.218 in the bracket of loan amount $30000 USD to $35000
3. Small Business which has average loan amount of around $14588 has higher charge-off's

Analysis for the loan characteristics(cont.)

3) Funded amount analysis with respect to loan status

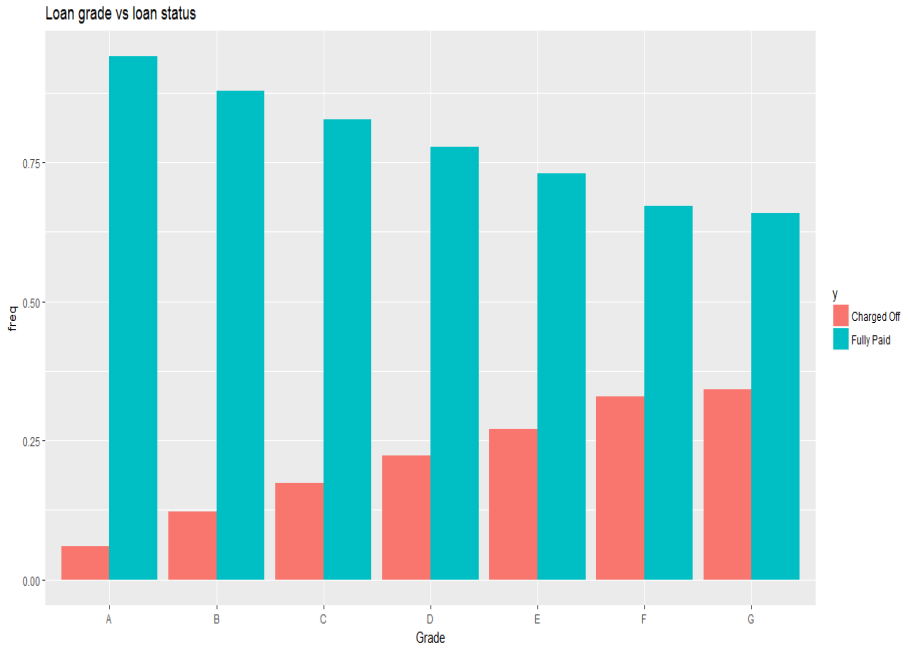


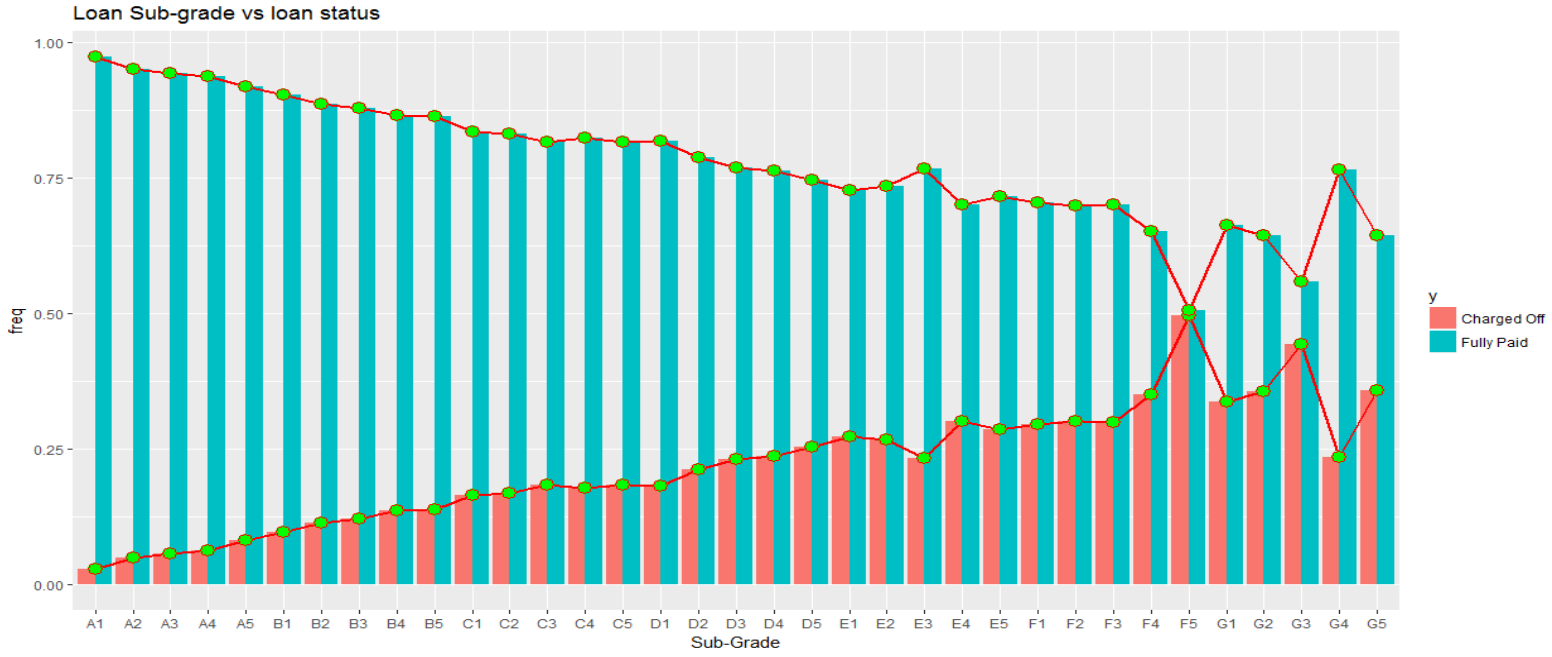
Findings

1. Probability of borrower defaulting is relatively higher (0.197) in the bracket of funded amount $30000 USD to $35000

Analysis for the loan characteristics(cont.)

4) Grade and Subgrade analysis with respect to loan status



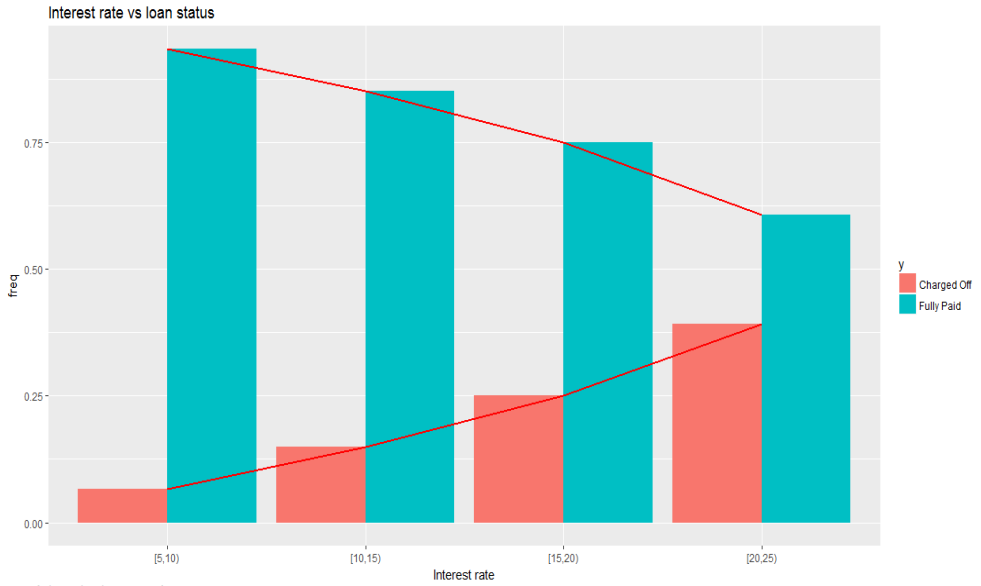


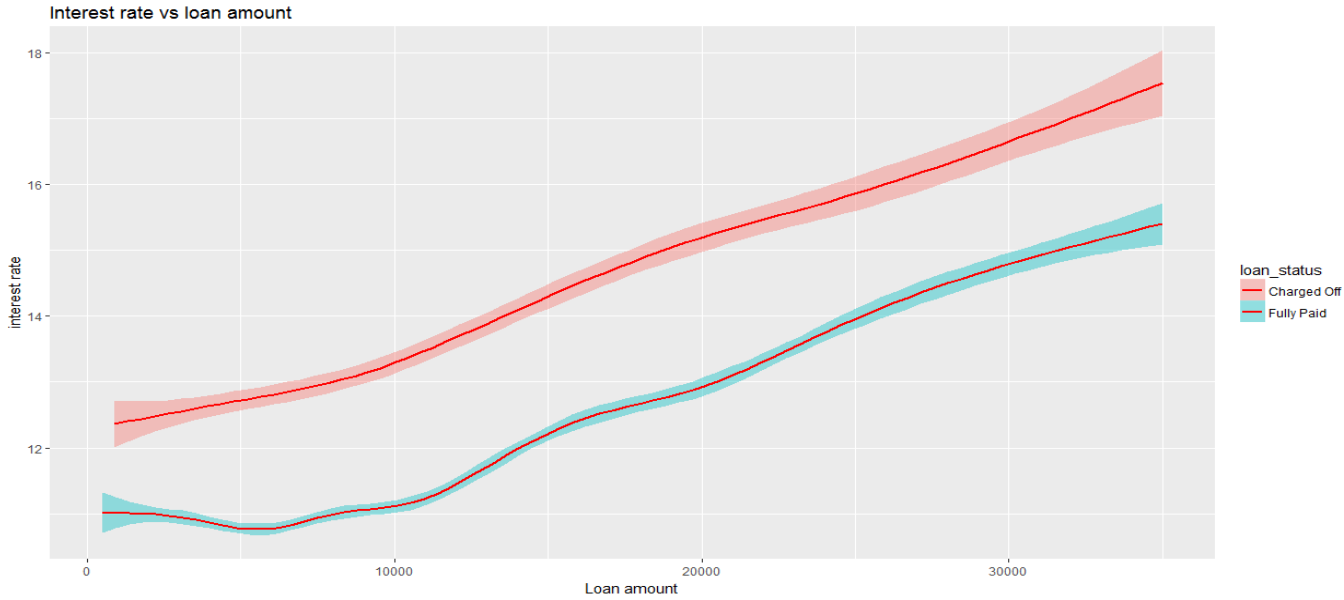
Findings

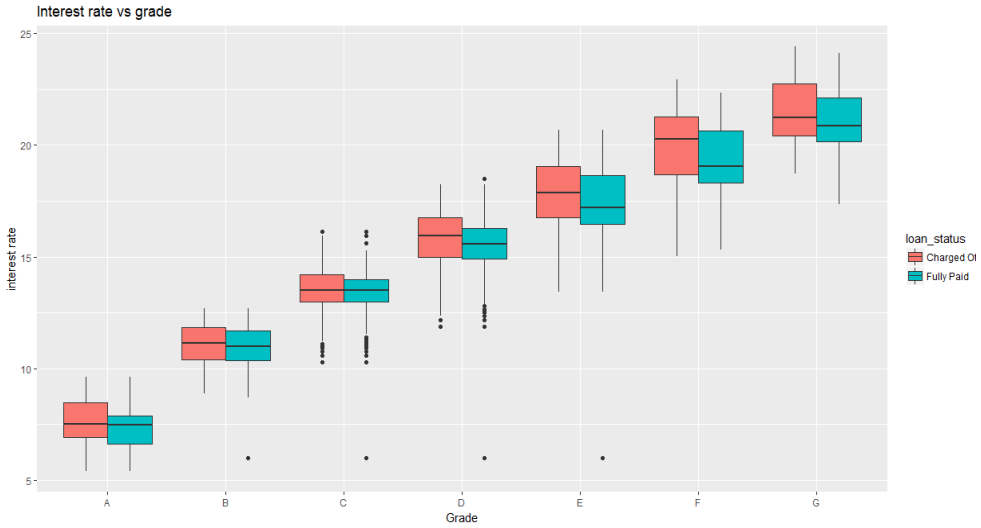
1. Probability of borrower defaulting increases as the grade lowers, the grade G shows the highest probability of default
2. Sub Grade F5 has 50% chances of a borrower defaulting
3. The plots show an increase in freq of charged off's for lower grades

Analysis for the loan characteristics(cont.)

5) Interest rate versus loan status





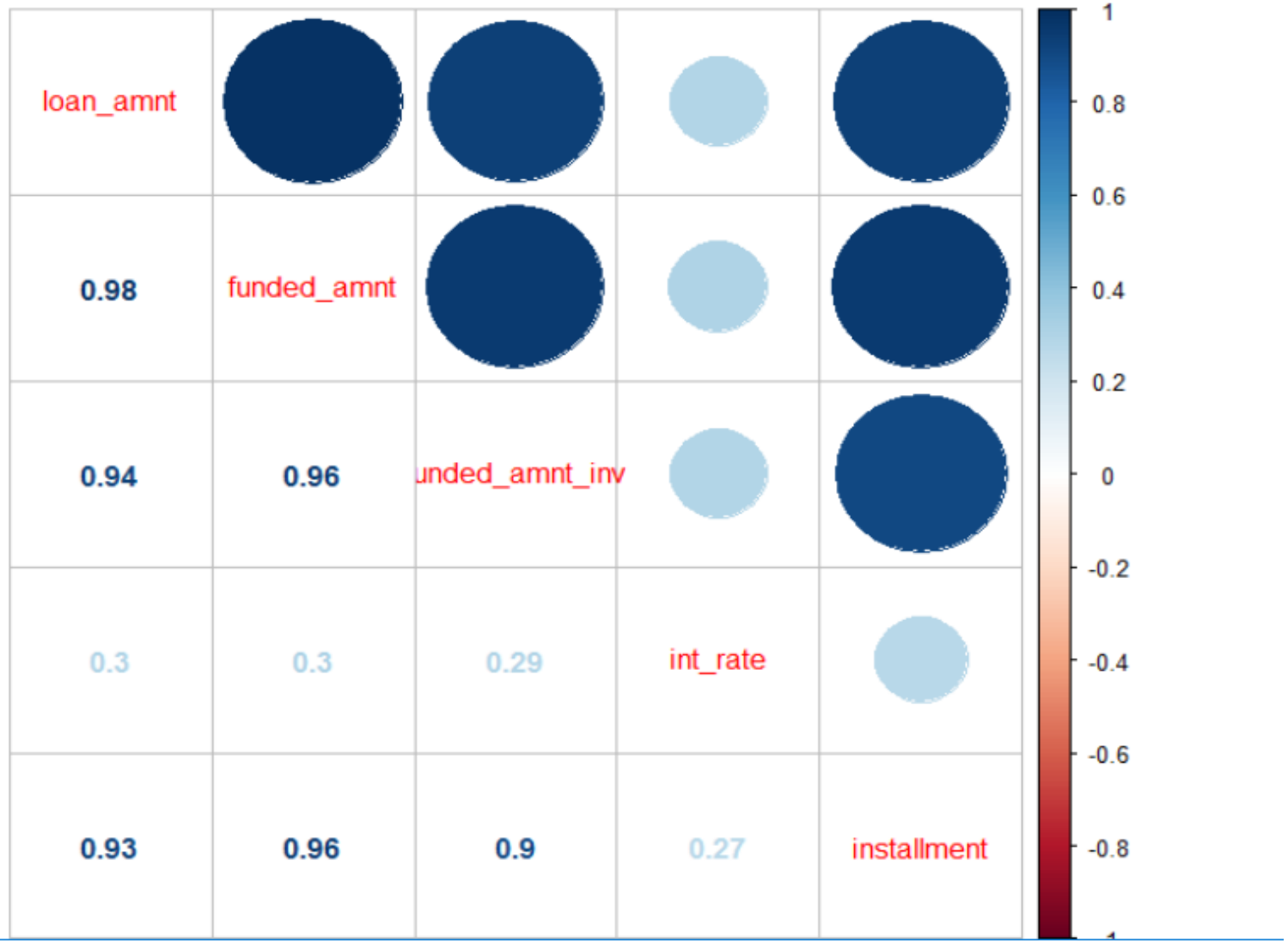


Findings

1. Loans with interest rate in the bracket of 20% to 25% have highest freq of default
2. Interest rates increases as the loan amount increases, higher interest rates have higher number of charge off's
3. Higher interest rates are assigned to lower grades

Analysis for the loan characteristics(cont.)

5) Pearson correlation plot for all the numeric variables in loan characteristics

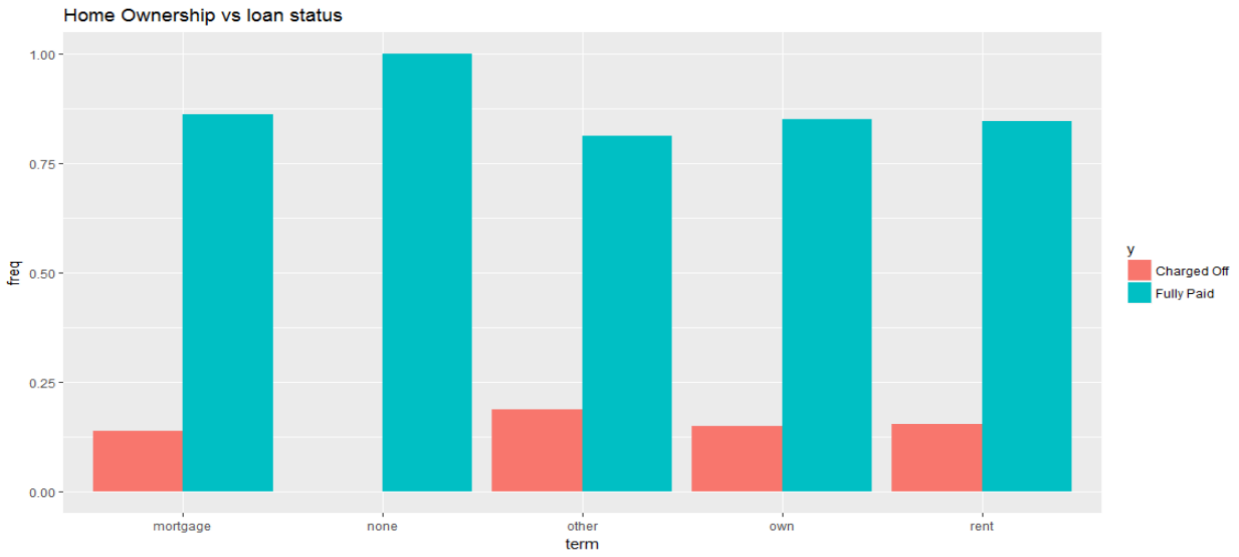


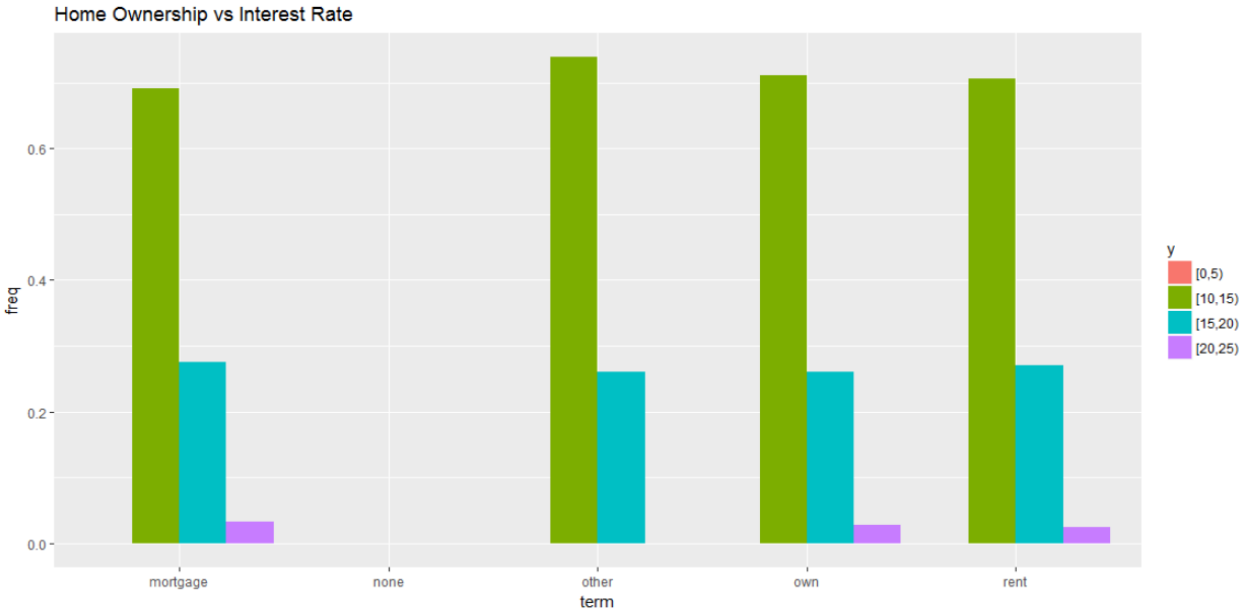
Findings

1. Numbers and the size of the circle are indicative of the strength of correlation among the variables.

Analysis for the borrower characteristics

1) Interest rate versus loan status





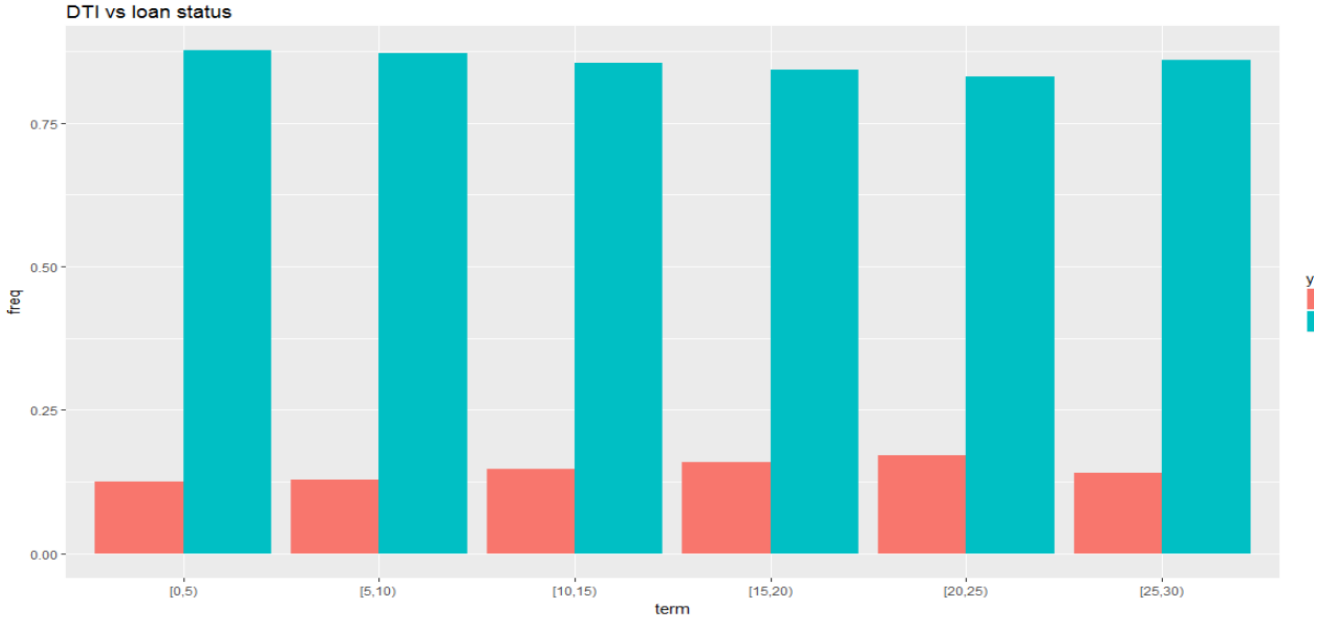
Findings

1. Borrowers with home ownership status as "other" have a higher default rate*
2. Borrowers with "other" as the home ownership status have the highest occurrences of interest rates in the bracket of (10%,15%)
3. People who are on mortgage have slightly higher chances of following into grade A

```
# x              A    C    D    E    F    G
  # mortgage 0.43 0.26 0.16 0.10 0.04 0.01
  # none     1.00 0.00 0.00 0.00 0.00 0.00
  # other    0.38 0.25 0.23 0.11 0.03 0.00
  # own      0.42 0.28 0.17 0.09 0.03 0.01
  # rent     0.32 0.32 0.22 0.10 0.03 0.01
"
```

Analysis for the borrower characteristics(cont.)

2) DTI versus loan status



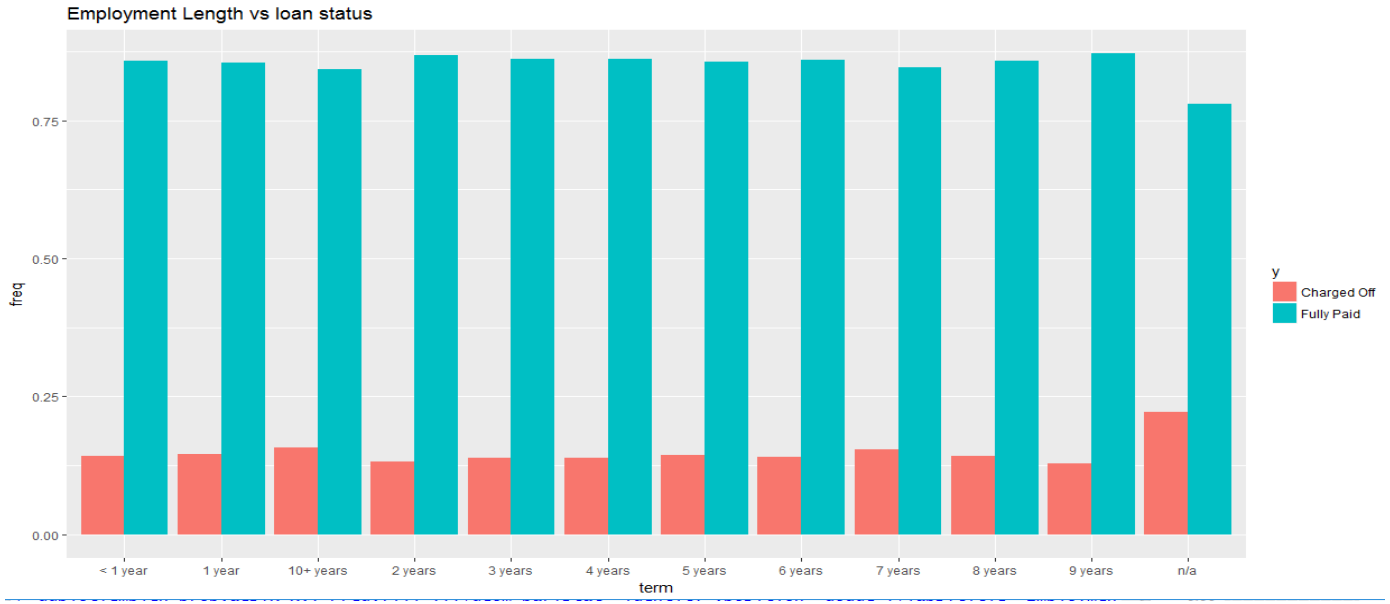
# x	Charged Off	Fully Paid
# [0,5)	0.124	0.876
# [5,10)	0.128	0.872
# [10,15)	0.146	0.854
# [15,20)	0.158	0.842
# [20,25)	0.170	0.830
# [25,30)	0.140	0.860

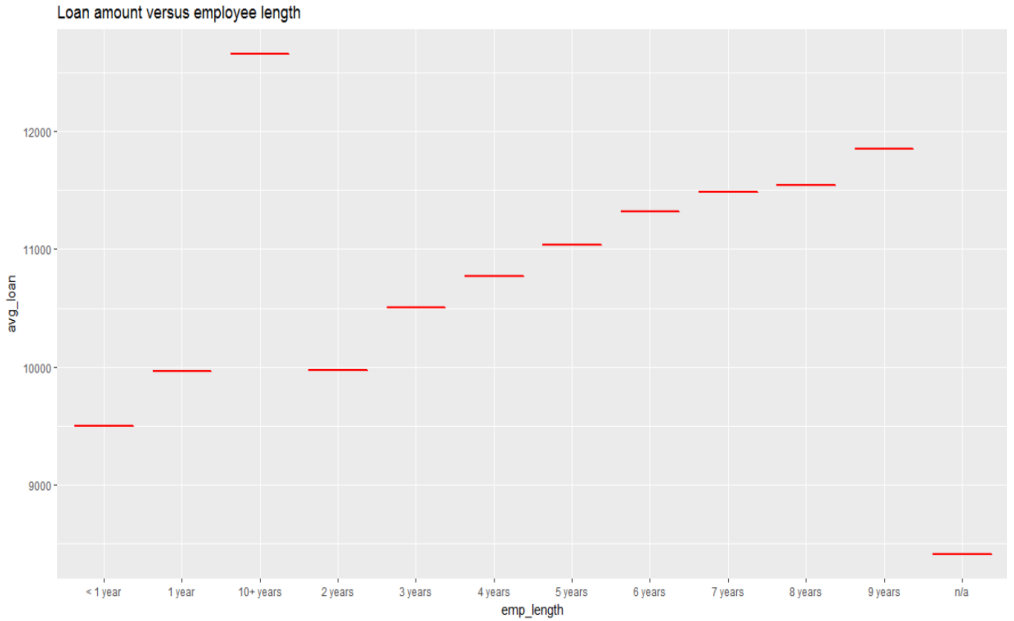
Findings

1. probability of a borrower defaulting is high(0.17) for borrowers with high dti

Analysis for the borrower characteristics(cont.)

3) Employee length (employee experience) versus loan status



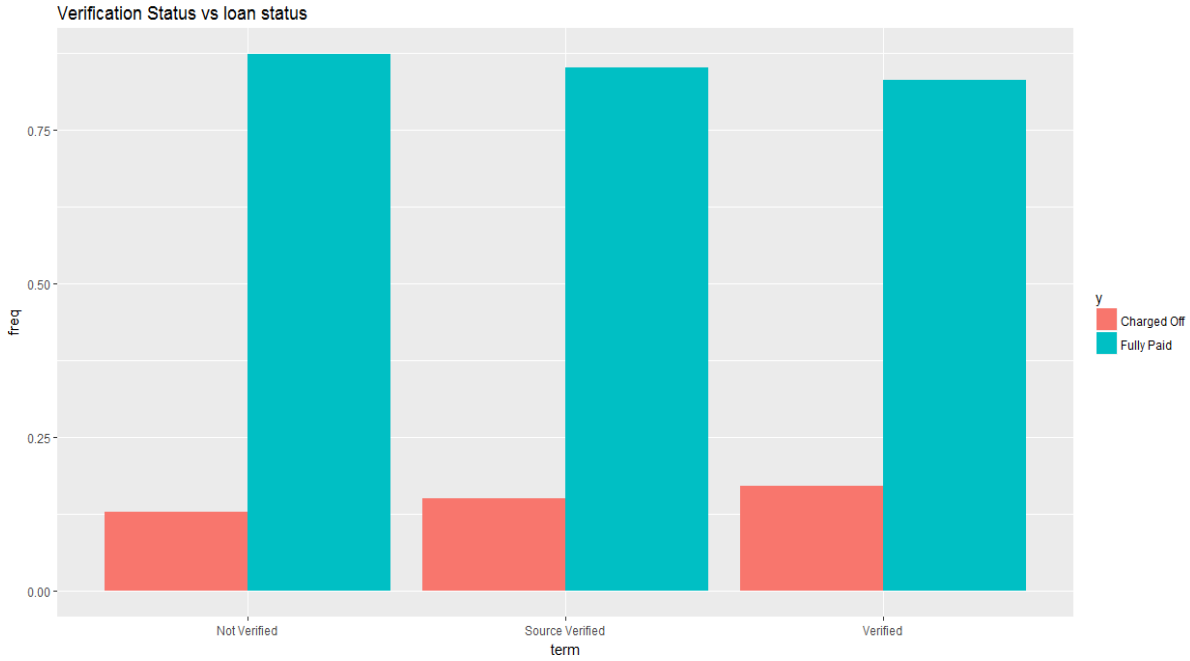


Findings

1. Employees with higher work exp. (10+ years) tend to default more than those with lower work exp.
2. One reason for borrowers employed longer defaulting more could be because they are taking larger loan amounts

Analysis for the borrower characteristics(cont.)

4) Verification status versus loan status



Grade	Not Verified	Verified
A	0.68	0.32
B	0.58	0.42
C	0.57	0.43
D	0.50	0.50
E	0.29	0.71
F	0.20	0.80
G	0.23	0.77

DTI	Not Verified	Verified
[0,10)	0.57	0.43
[10,20)	0.51	0.49
[20,30)	0.45	0.55

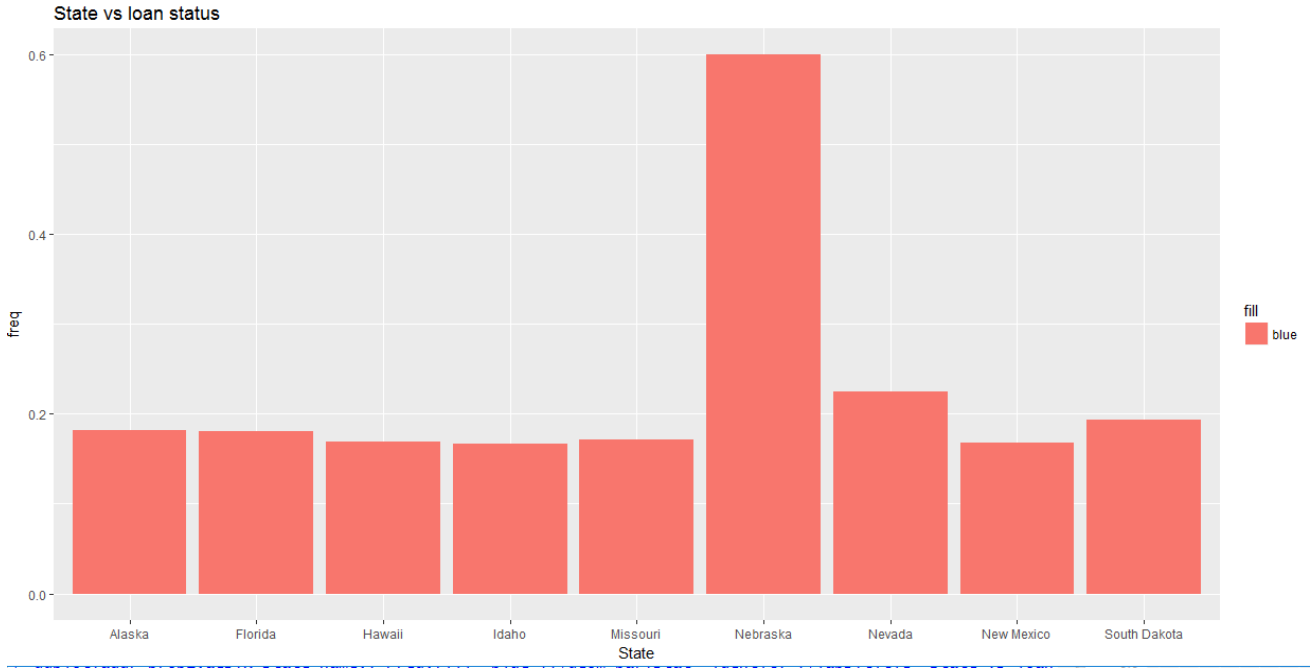
Findings

1. It's a common practice to verify income if the borrower shows a higher risk profile
2. More number of accounts of poorer grades are verified when compared to higher grades. % is higher from grade C and above
3. More number of accounts with higher dti tend to get verified 0.531 vs. 0.321 and 0.412

Analysis for the borrower characteristics(cont.)

5) addr_state: The impact of borrower's address

	x	y	Freq	state_name
1	NE	Charged Off	0.600	Nebraska
2	NV	Charged Off	0.225	Nevada
3	SD	Charged Off	0.194	South Dakota
4	AK	Charged Off	0.182	Alaska
5	FL	Charged Off	0.181	Florida
6	MO	Charged Off	0.171	Missouri
7	HI	Charged Off	0.169	Hawaii
8	NM	Charged Off	0.168	New Mexico
9	ID	Charged Off	0.167	Idaho



Findings

1. Nebraska followed by Nevada and South Dakota shows higher incidences of borrowers defaulting.