'Consumer Finance company' Case Study

(1 July 2018)

Prepared by,

Hariharan Subramaniam
Tuhin Roy
Sachin More
Karan Chavan

Problem Statement

A **consumer finance company** which specialises in lending various types of loans to urban customers faces two **types of risks** in terms making decisions :
- 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 Objective

To understand the **driving factors (or driver variables)** behind loan default, i.e. the variables which are strong indicators of default. So, The company can utilise this knowledge for its portfolio and risk assessment.

Problem solving methodology

Following steps will be taken ,

1.Create a metadata page ,detailing the basic information about the dataset.

2.Perform univariate and bivariate analysis on appropriate columns and identify patterns and relationships.

3.Plot graphs to demonstrate findings.

4.Summarise the finds in terms of hypothesis.

Meta Data

Meta Data	Description
Description	A consumer lending company's loan data along with a dictionary explaining the variables
Source	Upgrad
Format	loan.csv,Data_Dictionary.xlsx
Number of rows	39717
Number of columns	111
Each row is	Each row is an information of a customers loan record that includes customers details, loan transaction details etc
Subset	The dataset is a subset of Fullypaid, On-going ,Charged-off loans of customers who have been issued loans between the date range data range "2007-06-01" to "2011-12-01"
NAs	There are about 54 columns that entirely have NAs, these columns are ignored for analysis,Refer Appendix for the column names
0s	There are about 3 columns that entirely have 0s, these columns are ignored for analysis,Refer Appendix for the column names

Analysis

- **1.Loan status** - The dataset shows that there are 83% of the customers who are full paid,14.2% of the customers who are charged off and 2.9% of the customers who are current

- **2.Purpose** - From the 'loan purpose' perspective, most loans around 46.9% are taken for debt consolidation, second highest taken for credit card, third for other purposes and then Home improvement and so on.

  - **Purpose vs Loan status -** Most charged off customers are who have taken loan for **debt consolidation** I.e around 7%,second highest charged off customers are who have taken loan for 'other' purpose, third being for credit card and then for small business and so on.

- **3.Employment Length** - Customers who have employment length greater than 10 are the highest borrowers, followed by customers who have less than 1 year of experience and so on.There is 2.7% of the customers whose employment length is not known.

  - **Employment Length Vs Loan status -** 3.4 % of the loan that is charged off which is the highest in the charged off case are customers whose employment length is greater than **10 years,** this is followed by customers who have less than 1 year and so on.

  - **Average loan amount per Employment length vs loan status -** Average loan amount taken by customers who charged off are higher than the average loan amount taken by the full paid customer.Employment length 10+ years tops the list for charged off.

  - **Average annual income per Employment length vs loan status -** Average annual income of charged off customers are slightly lesser than the customers who fully paid and in current state.Employment length 10+ years tops the list for charged off.

Employment length Vs Loan Status


Employment length distribution
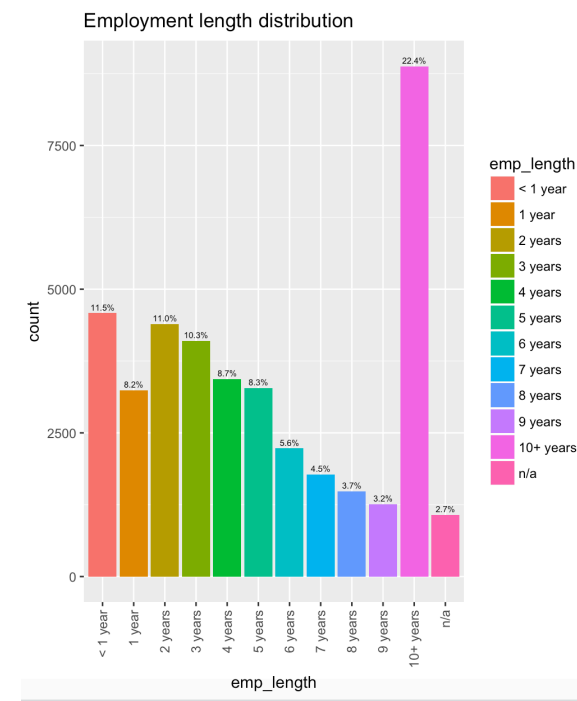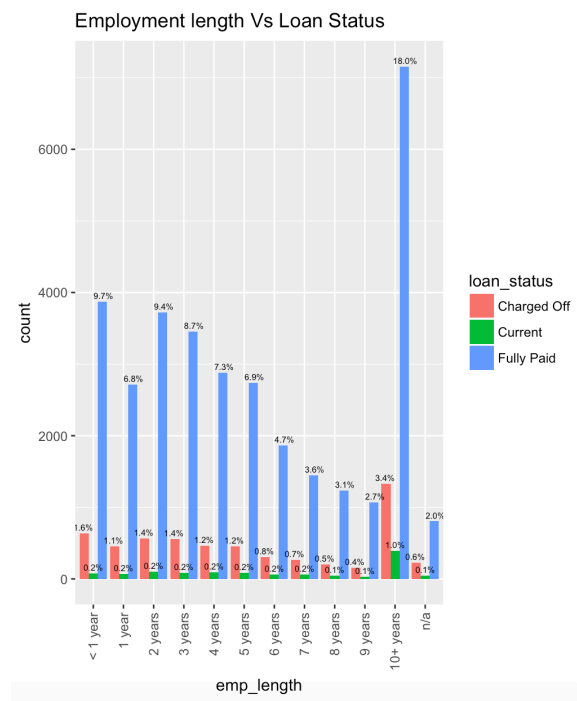
- **Grade** - Grade 'B' shows the highest distribution followed by Grade 'A' then C and so on.
  - **Grade Vs Loan status** - More charged off customers are in the grade 'B' (3.6%)then in 'C'(3.4%),'D'(2.8%),'E'(1.8%),'A' (1.5%)and os on
  - **Grade per employment length Vs Loan status -** More charged off customers are observed for employment length greater than 10 years
  -

Overall distribution of Grade

Grade Vs Loan Status

Analysis
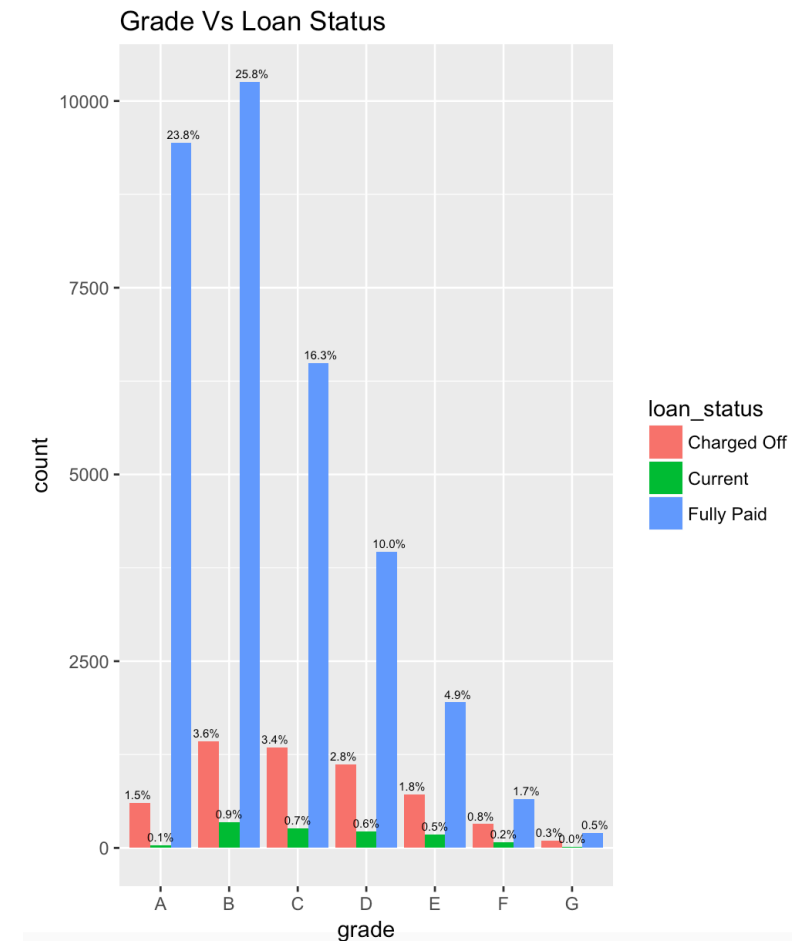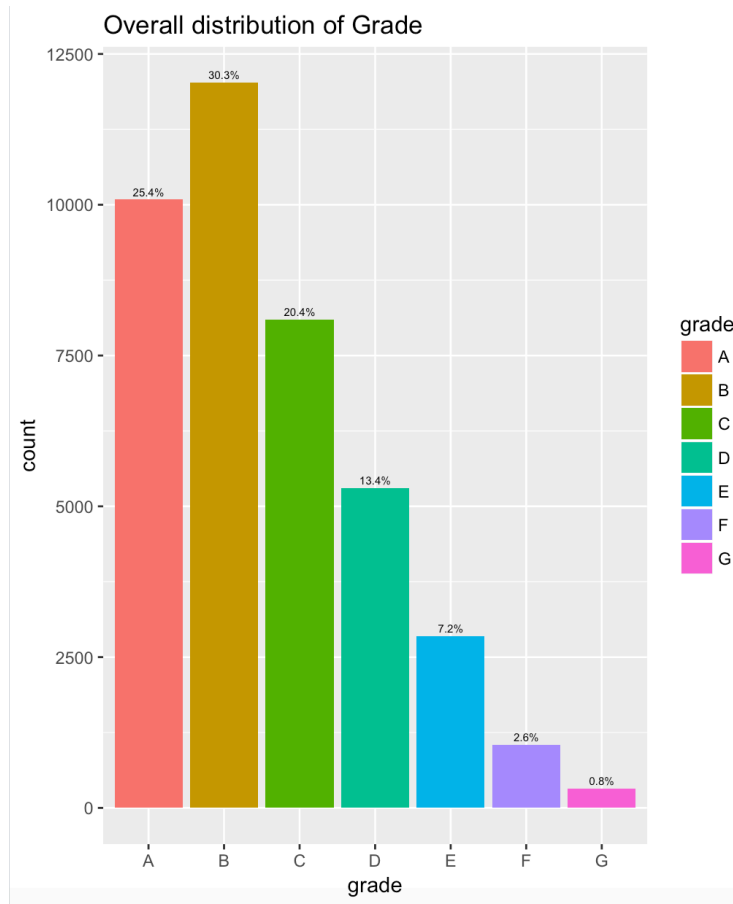
- **Home Ownership** - Overall distribution of 'Home Owership' shows that about 47.6% of the customers stay on a rented home,44.5% of customers stay on a mortgage home, only 7.7% of the customers stay in their own home.0.2% of the customers home ownership is not known.
  - **Home Ownership Vs Loan status** - highest number of charged off customers are who stay in a rented home I.e about 7.1% ,the second highest being who stay in mortgaged home i.e about 5.9%.
  - **Home Ownership Vs loan status per employment length** - High number of charged off customers are whose employment length is greater than 10 years and who live in mortgage homes followed by customers who live in rented home. However cumulatively customers staying in rented home show higher charged off rate compared to who live in portage home across all employment length.
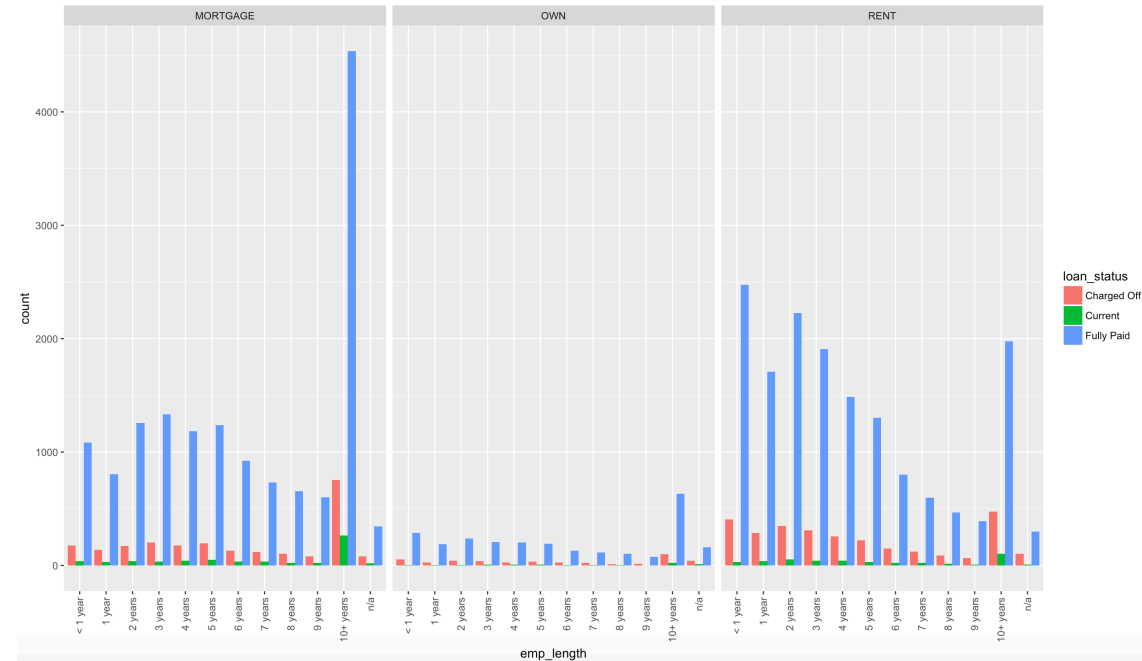

Overall distribution of Home Ownership
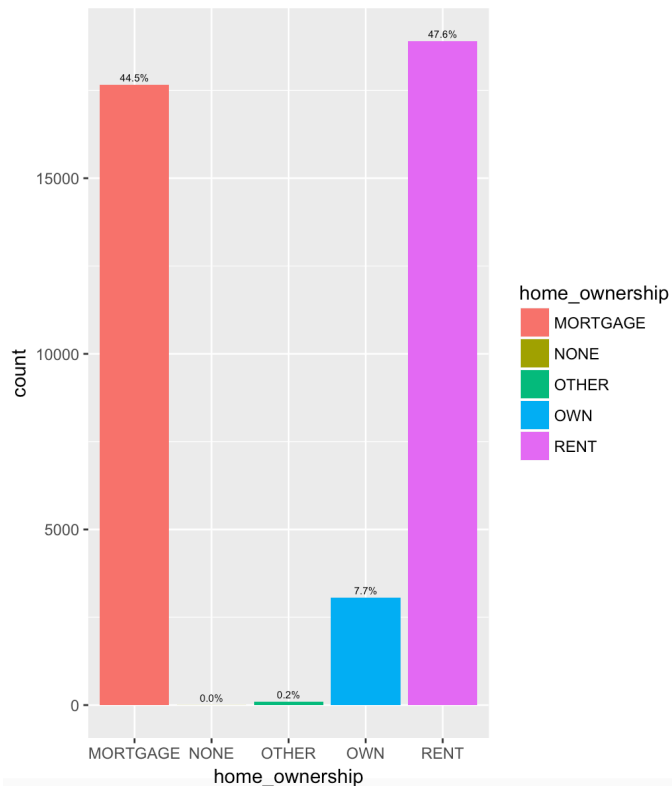

Home Ownership Vs loan status

Home Ownership Vs Loan status per Employment length

Analysis
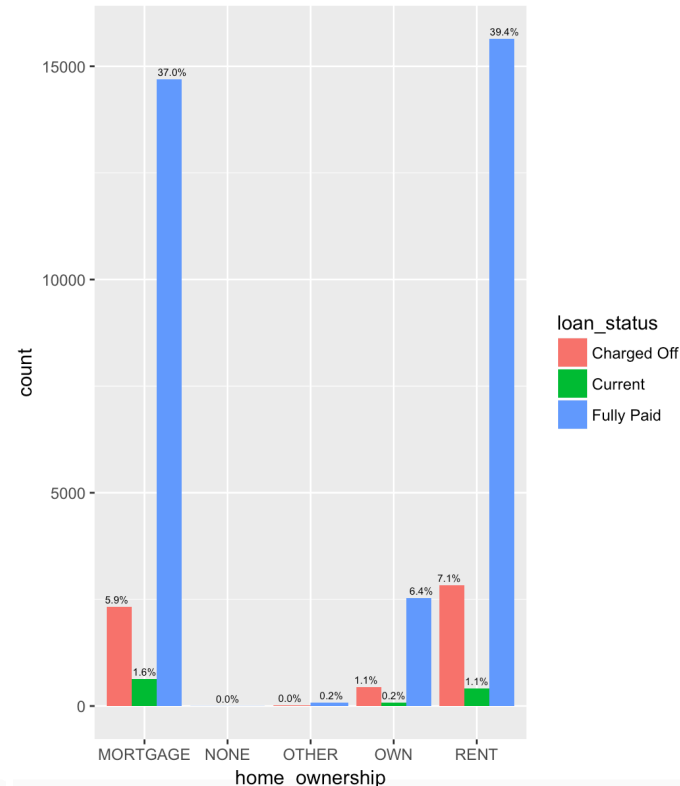
- **1.Number of public record bankruptcies' -** When plotting a relationship between 'Number of public record bankruptcies' filed by people with various employment length against the verification status.It is observed that most of the customers who have higher number of 'public record bankruptcies' are who have employment length greater than 10 years.And people with employment length greater than 10 years with 'Not Verified' are at peak

- **2**.**Purpose** -

  - **Car** - People whose average income less than 150000 have taken Car loan and a proportion of people whose employment length is of **3 and 5** years are charged off.

  - **Credit Card** - People whose average income less than 200000 have taken Credit card loan and a proportion of people whose employment length is other than **1,5,8,9** years are charged off.

  - **Debt_consolidation** - People whose average income less than 200000 have taken Debt_consolidation loan and a proportion of people across **all** employment are charged off.

  - **Home Improvement** - People with average income less than 300000 have taken home improvement loan and a proportion of people across employment lengths **1,35,8,10+** years are charged off.

  - **House** - People with average income less than 100000 have taken loan for house purpose and a proportion of people across employment lengths **2 years** are charged off and people with employment length 7years are fully charged off

  - **Major purchase** -People with average income less than 175000 have taken loan for major purchase and a proportion of people across employment lengths **9,10+,less than 1** year are charged off .

  - **Medical** - People with average income less than 150000 have taken loan for Medical and a proportion of people across employment lengths **less than 1 year** and **3,5,10+** years are charged off

  - **Moving** - People with average income less than 150000 have taken loan for moving purpose and a proportion of people across employment lengths **3 and 4 years** are charged off

  - **Other** - People with average income less than 200000 have taken loan for other purpose and people with **less that 1 year** have higher charged off rate relative to proportion of people across other employment lengths .

  - **Renewable Energy** - **None** from renewable energy are charged off.

  - **Small Business** - People with average income less than 200000 have taken loan for small business purpose and a proportion of people across all employment lengths are charged off and people with employment length **8years** are fully charged off

  - **Vacation** - People with average income less than 160000 have taken loan for vacation and a proportion of people across employment lengths **3,4,10+** years are charged off

  - **Wedding** - People with average income less than 100000 have taken loan for wedding purpose and a proportion of people across employment lengths **2** years are majorly charged off compared to people in 10+ years .

Plot for purpose

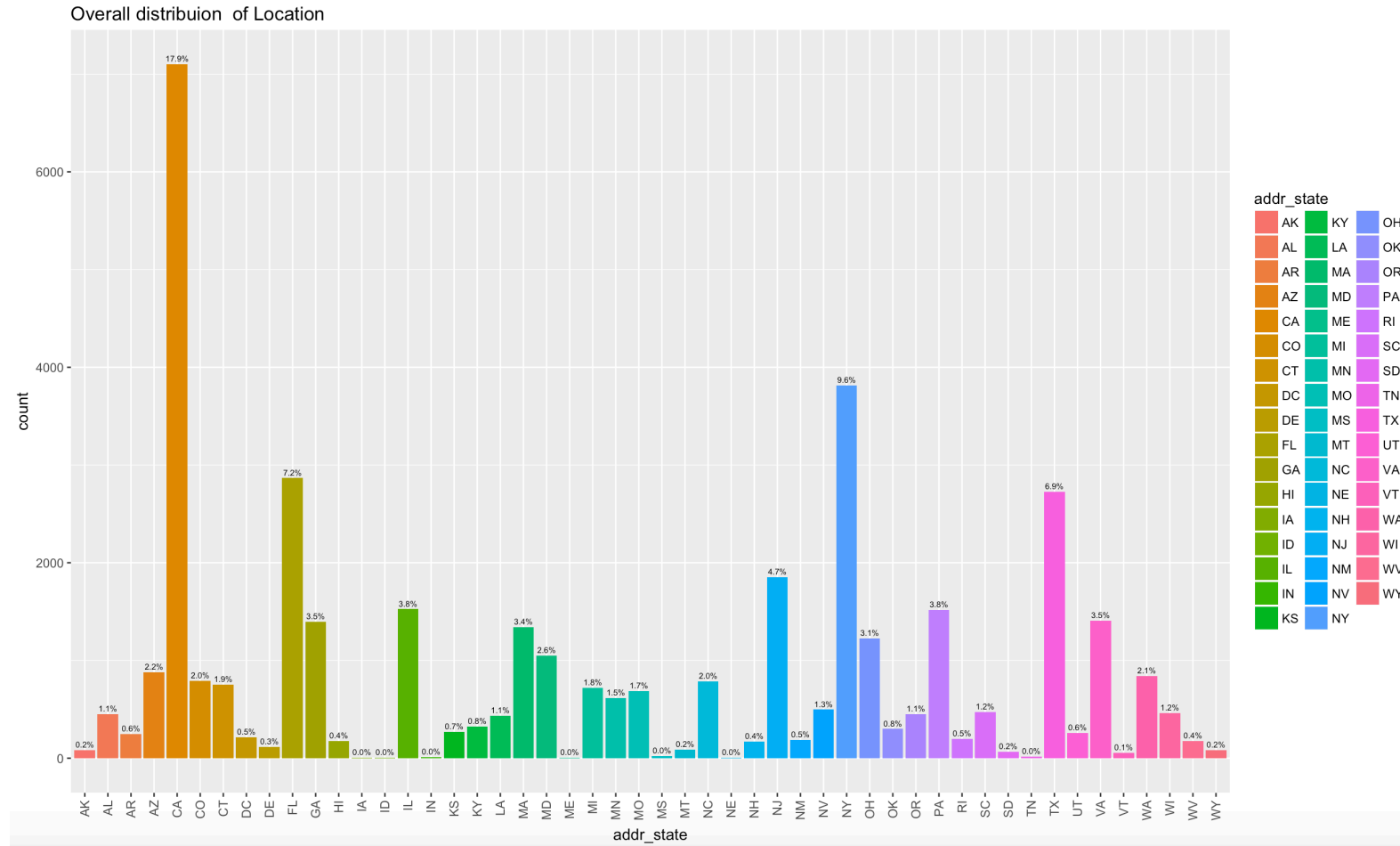Purpose Vs Loan status per Employment length

Analysis
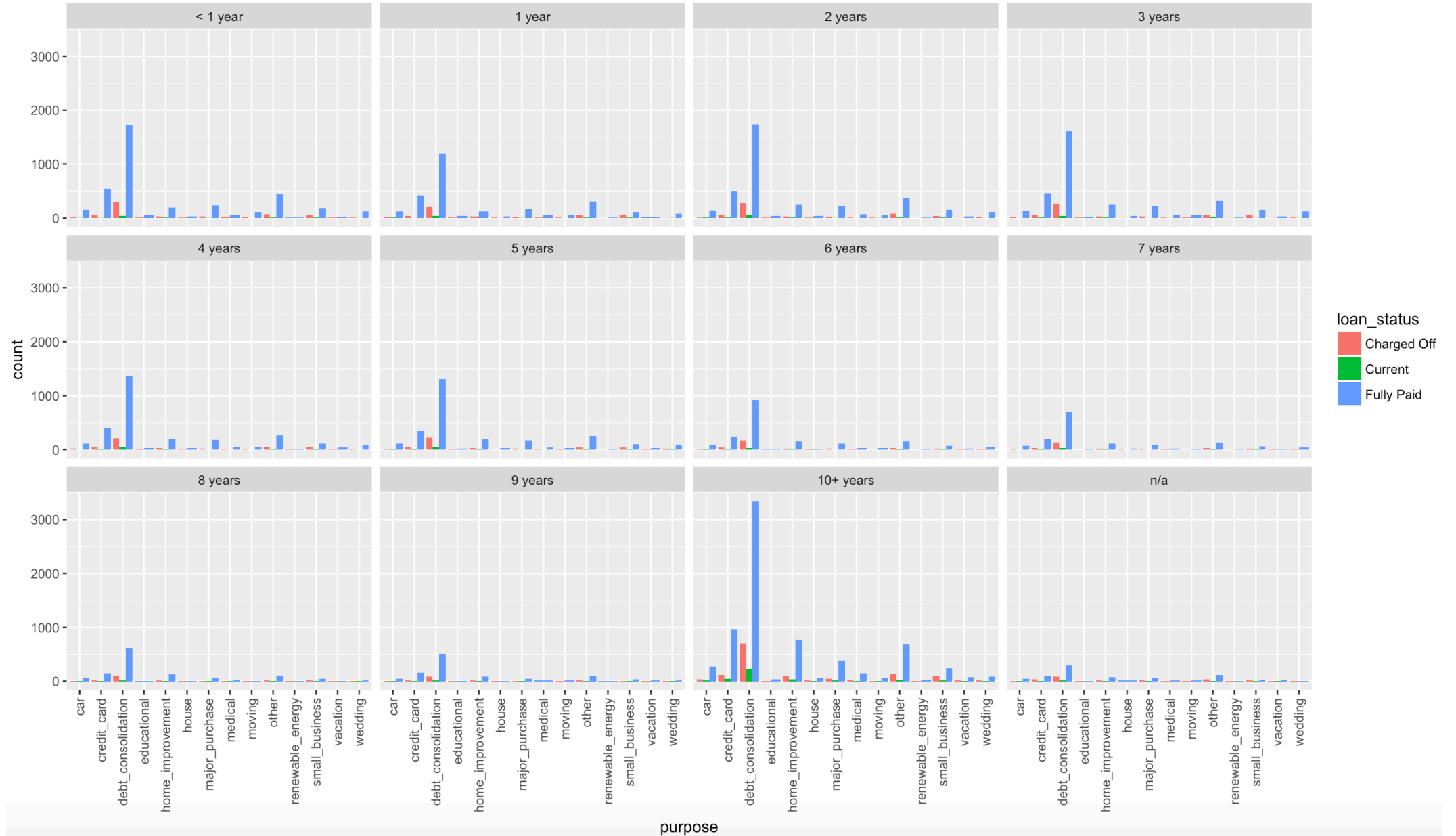
- **Verification status** - About 42.6% of the Customers are Not verified
  - **Verification status Vs loan status**, customers who are 'not verified' and just 'verified' have higher proportion for charge off cases.
  - **Verification status Vs loan status per employment length**, plot shows customers whose employment length is less than 1 year and who are not verified have higher charged off rate, however employees whose employment length is greater than 10 years and are verified but not source verified have higher charge off rate
 -
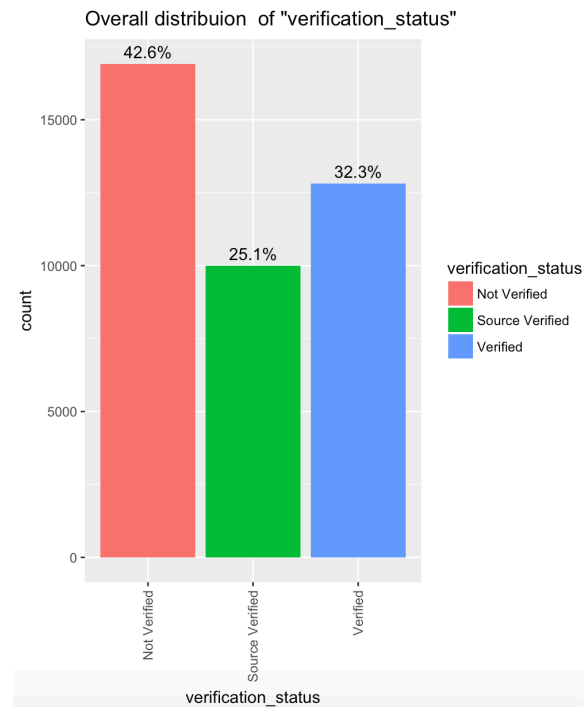
- **Term** - It is understood that the lending company initially was providing loan for a term of only 36 months from the year 2007 - 2010.It was in May 2010 the company started offering loans for terms 60 months.When the term length was analysed against loan status ,it is understood the proportion of customers 'Charged off' whose term is 60 months are relatively higher than the proportion of customers 'Charged off' paying loans on 36 months term
    - **Overall distribution of Term**,About 73.3% of the customers have opted for 36 months and 26.7 % have opted for 60 months
    - **Term Vs loan status,**About 23% of the customers who opted for 60 months term are charged off which is higher than the customers who opted for 36 months which is about 11%
    - **Term Vs loan status per Employment length**, customers whose employment length is greater than 10 years are highly charged off when their term for payment is 60 months , the second highest is when employment length is greater than 10 years term is 36 and the third being employment length less than 1 year and term 36 months

Analysis
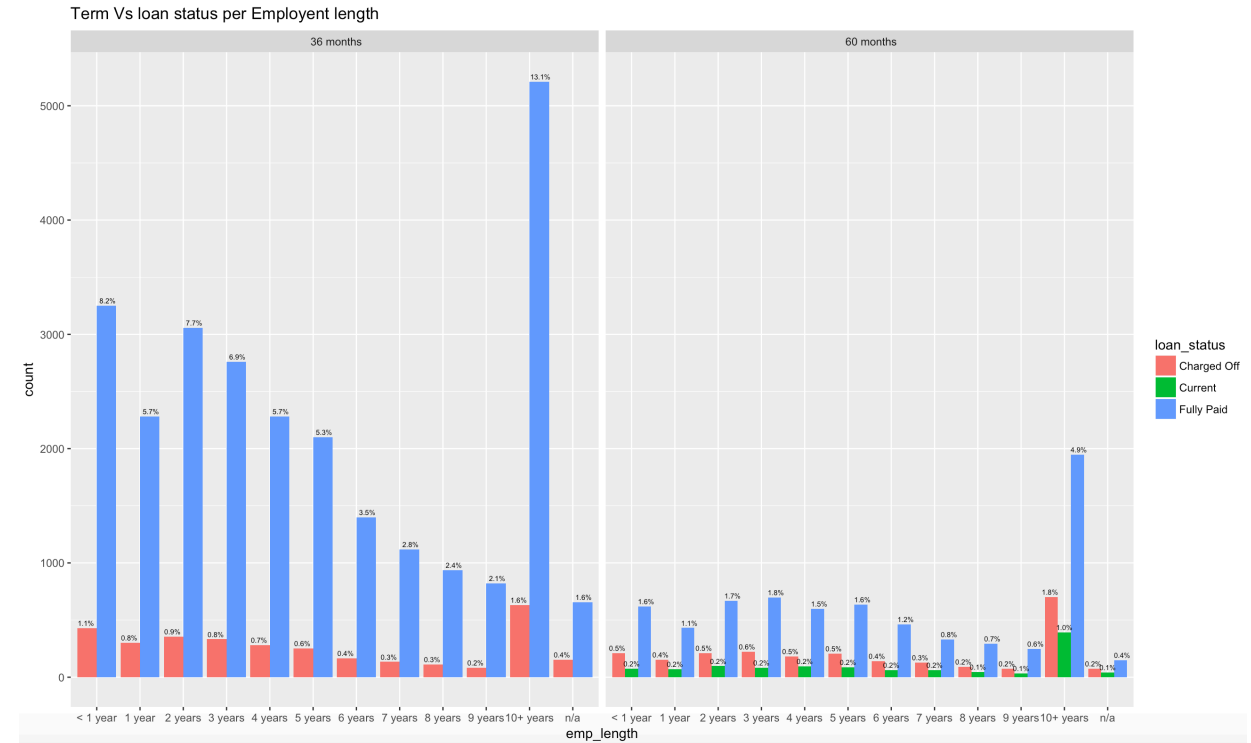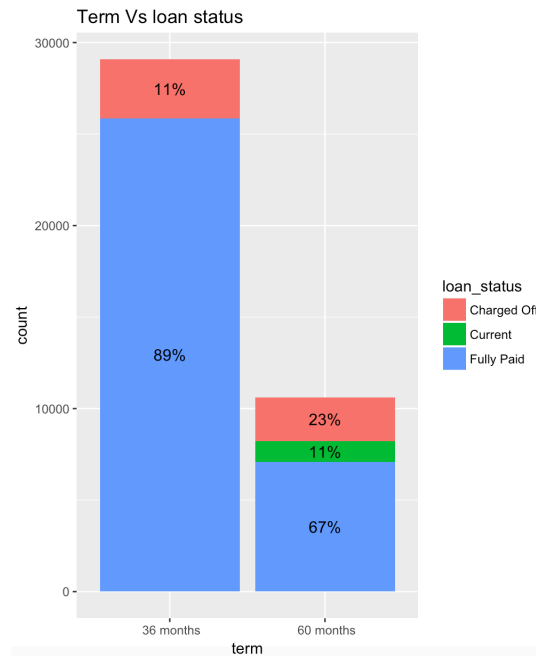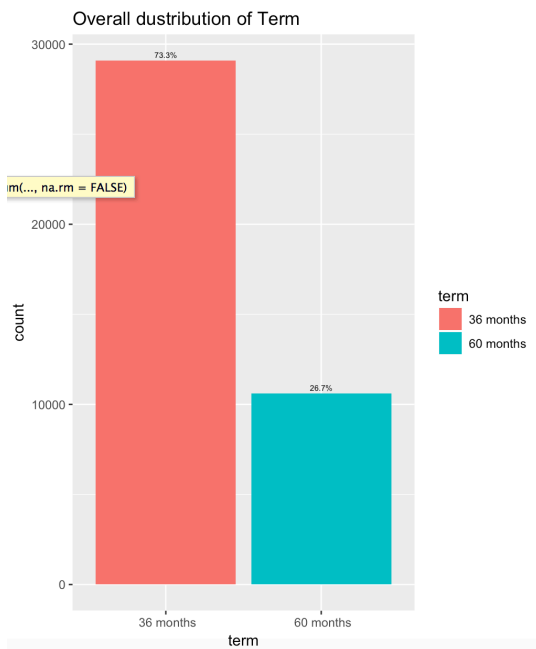
- **Location** - Overall distribution , California has the highest number of customers

- Location VS Charged off,California followed by Florida and then Newyork have higher charged off rate

Overall distribuion of Location

Correlation between all numeric variables
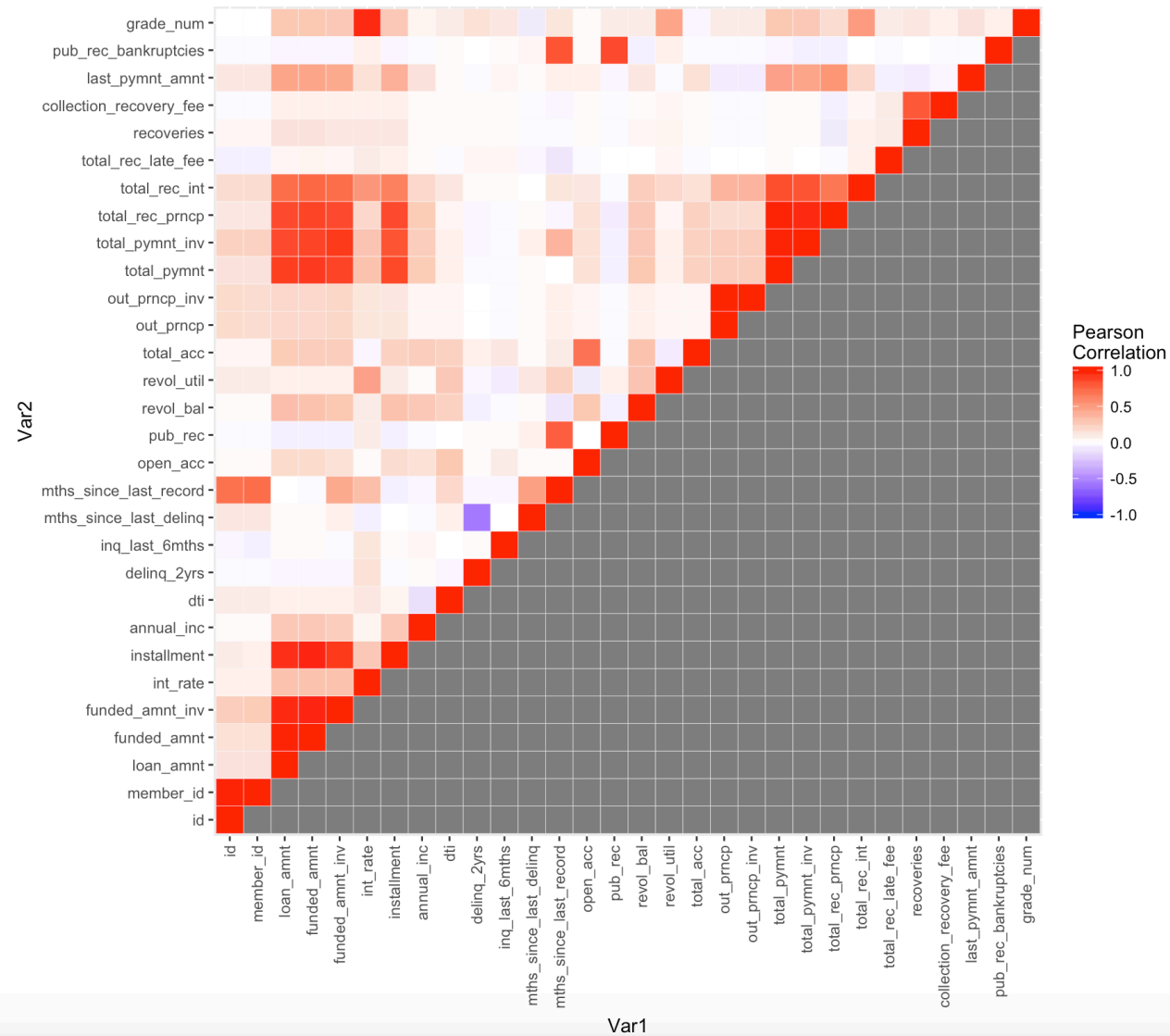
Conclusion

- Variables such as Purpose, employment length,Home ownership,Term,Verification status,Grade,Location were key parameters for analysis.

- Customer whose employment length greater than 10 years and less than 1 year are highly likely to default.

- Customers opting for Debt consolidation loan are highly likely to default

- Customers living in homes that are mortgaged and rented are highly likely to default

- Customers whose grade are 'B','C','D' are highly likely to default

- Customers living in California,Florida and newyork are highly likely to default

Columns that entirely have NAs

- There are about 54 columns that entirely have NAs as listed below
- [1] "mths_since_last_major_derog","annual_inc_joint" ,"dti_joint"
- [4] "verification_status_joint","tot_coll_amt","tot_cur_bal"
- [7] "open_acc_6m","open_il_6m","open_il_12m"
- [10] "open_il_24m","mths_since_rcnt_il","total_bal_il"
- [13] "il_util","open_rv_12m","open_rv_24m"
- [16] "max_bal_bc","all_util","total_rev_hi_lim"
- [19] "inq_fi","total_cu_tl","inq_last_12m"
- [22] "acc_open_past_24mths","avg_cur_bal","bc_open_to_buy"
- [25] "bc_util","mo_sin_old_il_acct","mo_sin_old_rev_tl_op"
- [28] "mo_sin_rcnt_rev_tl_op","mo_sin_rcnt_tl","mort_acc"
- [31] "mths_since_recent_bc","mths_since_recent_bc_dlq","mths_since_recent_inq"
- [34] "mths_since_recent_revol_delinq" ,"num_accts_ever_120_pd","num_actv_bc_tl"
- [37] "num_actv_rev_tl","num_bc_sats","num_bc_tl"
- [40] "num_il_tl","num_op_rev_tl","num_rev_accts"
- [43] "num_rev_tl_bal_gt_0","num_sats","num_tl_120dpd_2m"
- [46] "num_tl_30dpd","num_tl_90g_dpd_24m","num_tl_op_past_12m"
- [49] "pct_tl_nvr_dlq","percent_bc_gt_75","tot_hi_cred_lim"
- [52] "total_bal_ex_mort","total_bc_limit","total_il_high_credit_limit"

Columns that entirely have 0s

- There are about 3 columns that entirely have 0s
- [1] "collections_12_mths_ex_med" , "chargeoff_within_12_mths" , "tax_liens"