Gramener Case Study

Group Name:

1. Vijay Choudhary
2. Aman Chaudhary
3. Abhinaw Kumar
4. Shwetha S

Problem Statement and Business Objectives

Problem Statement

- To understand how consumer attributes and loan attributes influence the tendency of the loan request to be 'defaulted' or not. 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.

Business Objectives

- 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

Business Constraints:

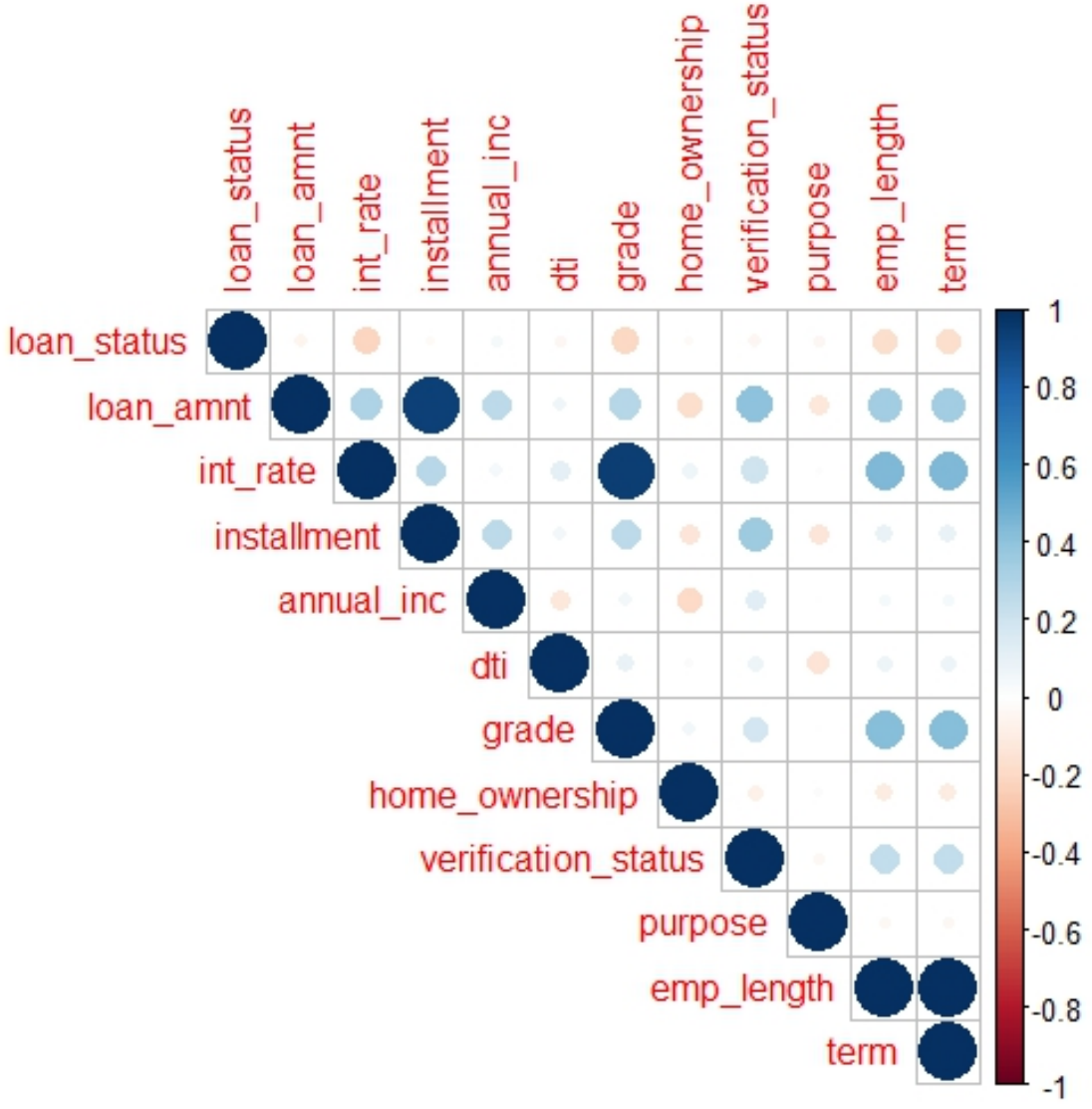
- Consider only consumers whose loan application is approved in analysis.

Data - Unique 39717 records for past loan applicants (2007-2011)

Data

- 39717 unique records with 111 observations including customer demographic , customer behavioral and loan application related attributes.
- 3 types of Loan Status (**Fully paid**, **charged Off** , **Current**)
- It contains the records for all the loans issued through the time period 2007 to 2011
- For this assignment, we considered the loan accepted related data only not the rejected one.

Data - Correlations between different attributes



Positive Correlation :

- Loan Amount and Installment
- Interest rate and grade
- Emp Length and term

Negative Correlation :

- Loan Status and Interest Rate
- Loan Status and Grade
- Loan Status and Emp Length

EDA - Problem solving methodology

Data Collection – Loan dataset given by the company

Data Cleaning –

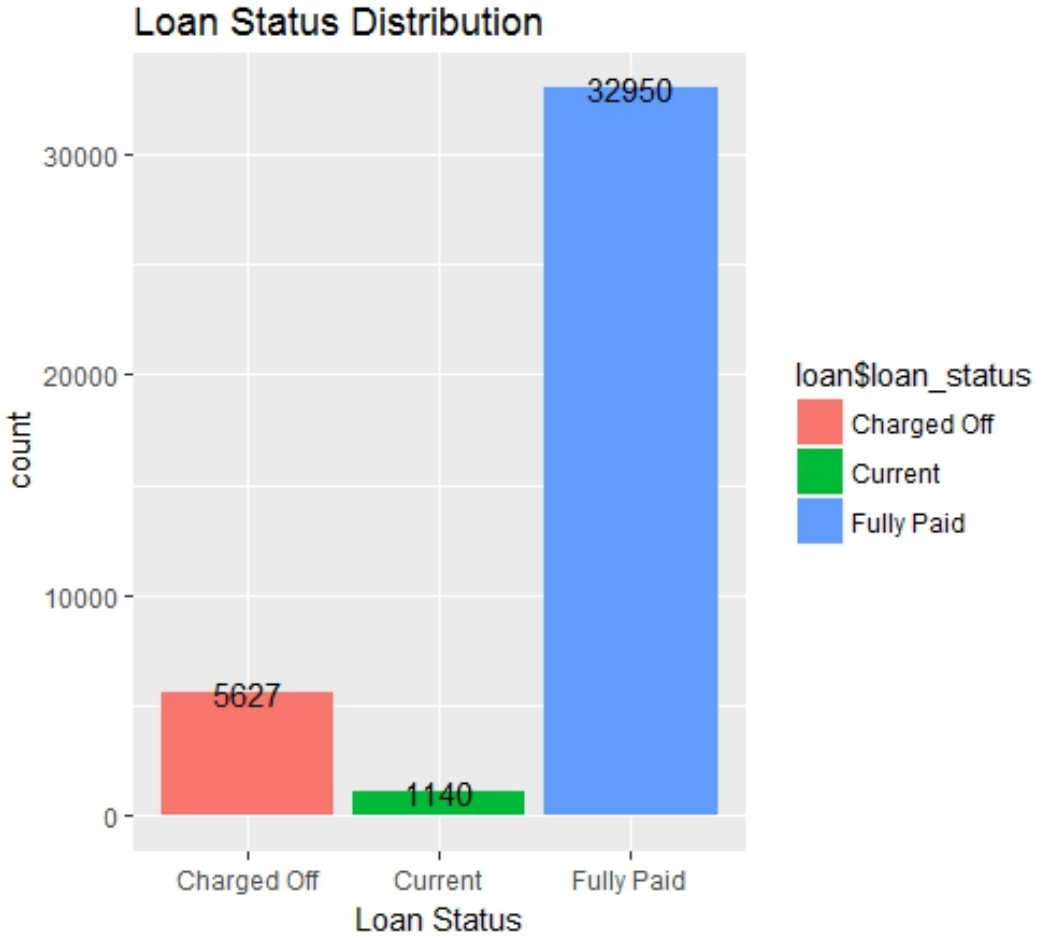
- Finding percentage of NA values in the columns like emp_length and revol_util and removing them
- Removing unnecessary non-numeric values from the numeric data column like removing % sign from int_rate and revol_util columns and the literal 'year' and 'month' from emp_length and term column respectively

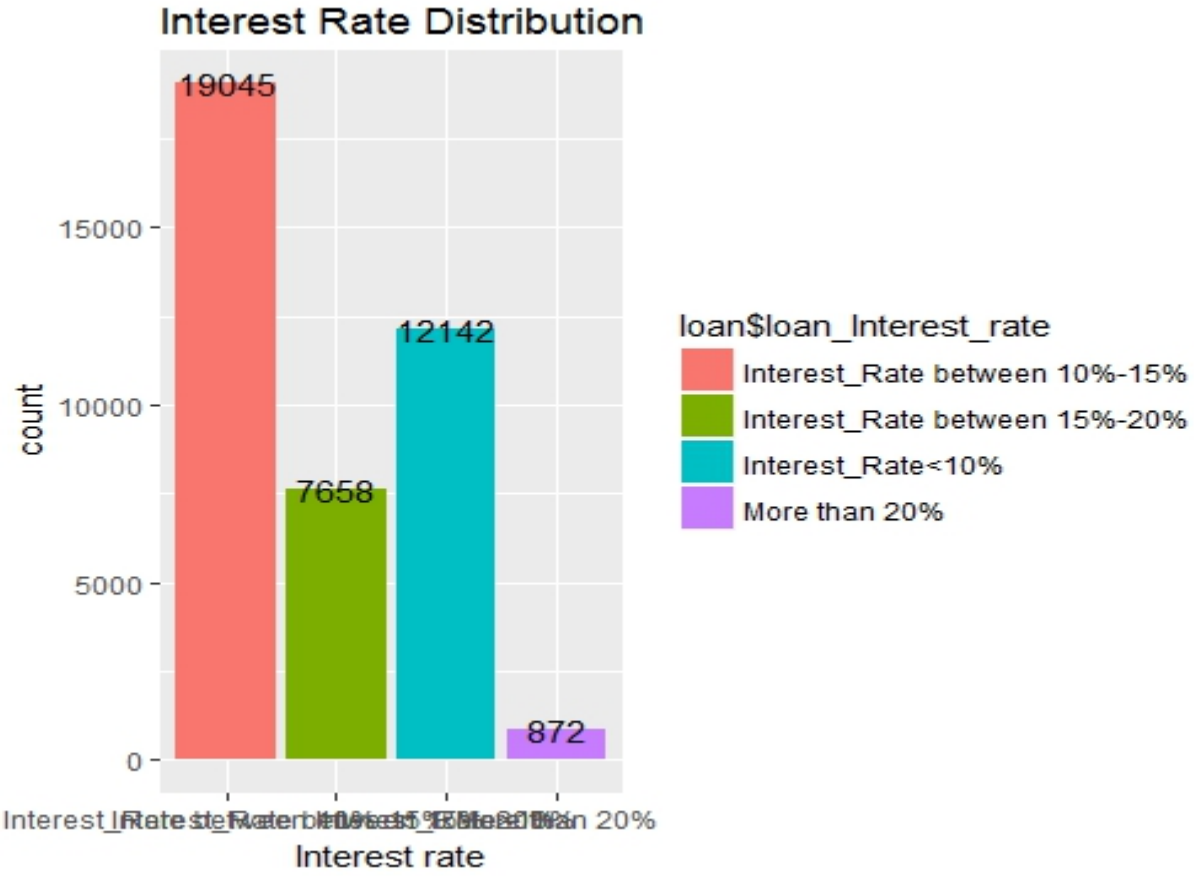
Univariate Analysis – Analyzing the distribution of each variables individually

Multivariate Analysis – Finding the relationship between each variables

Conclusion – Finding which variables drives loan defaulting

Univariate Analysis – Loan status and Interest Rate Analysis

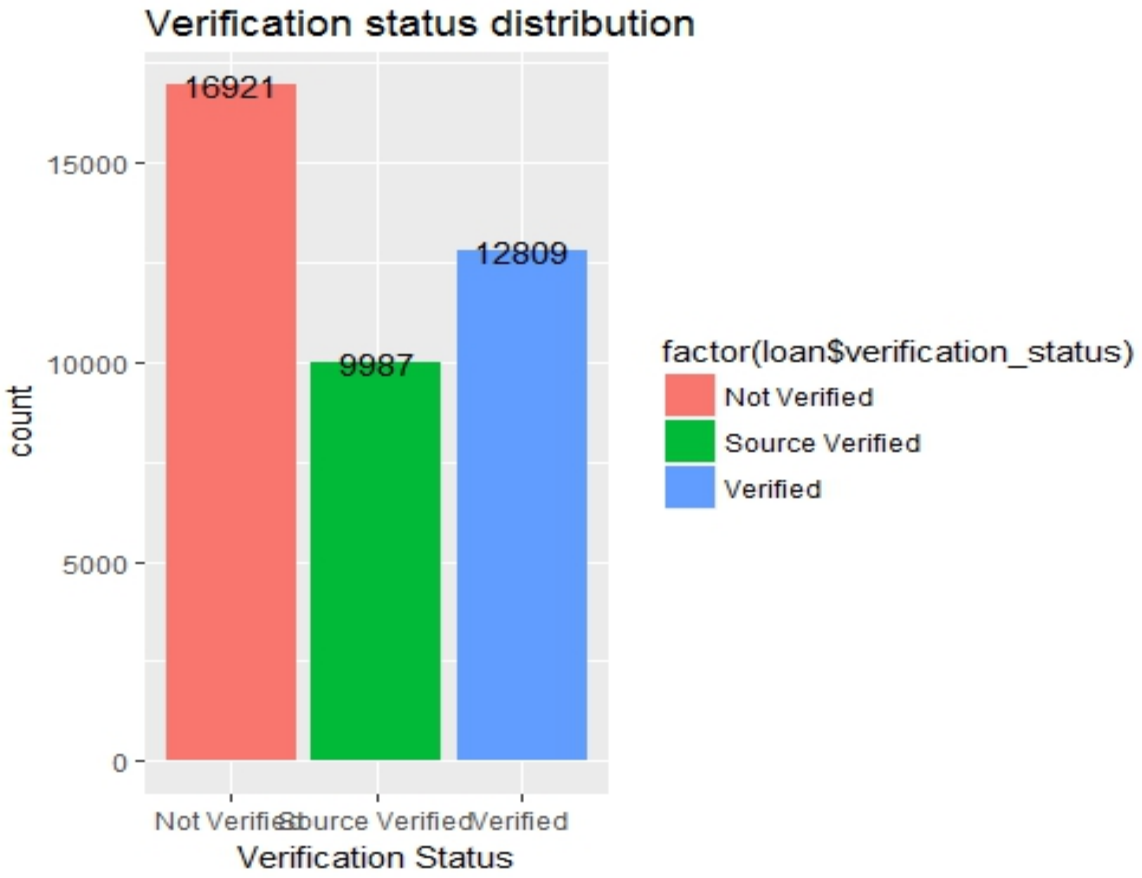


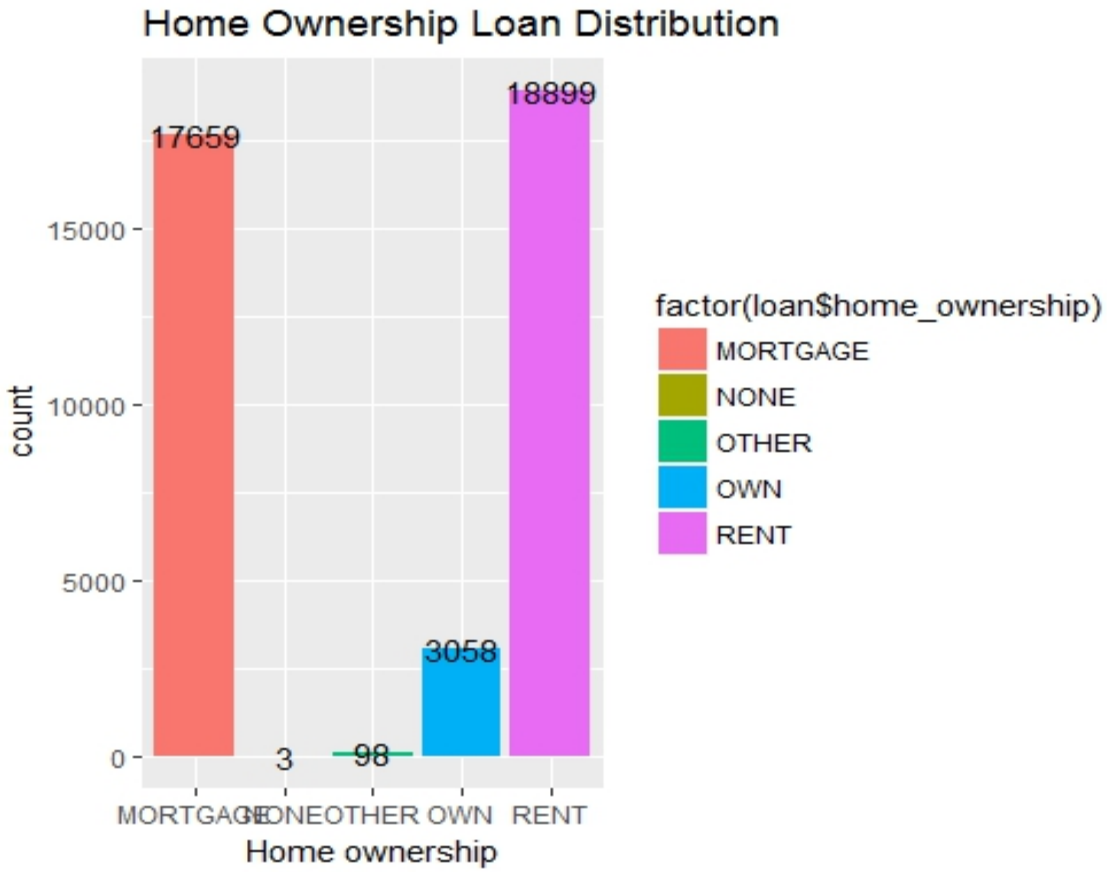


14 % of the Loans have been Charged Off

Most of the Loans have interest rate Between 10%-20%

Univariate Analysis – Verification and Home ownership Analysis

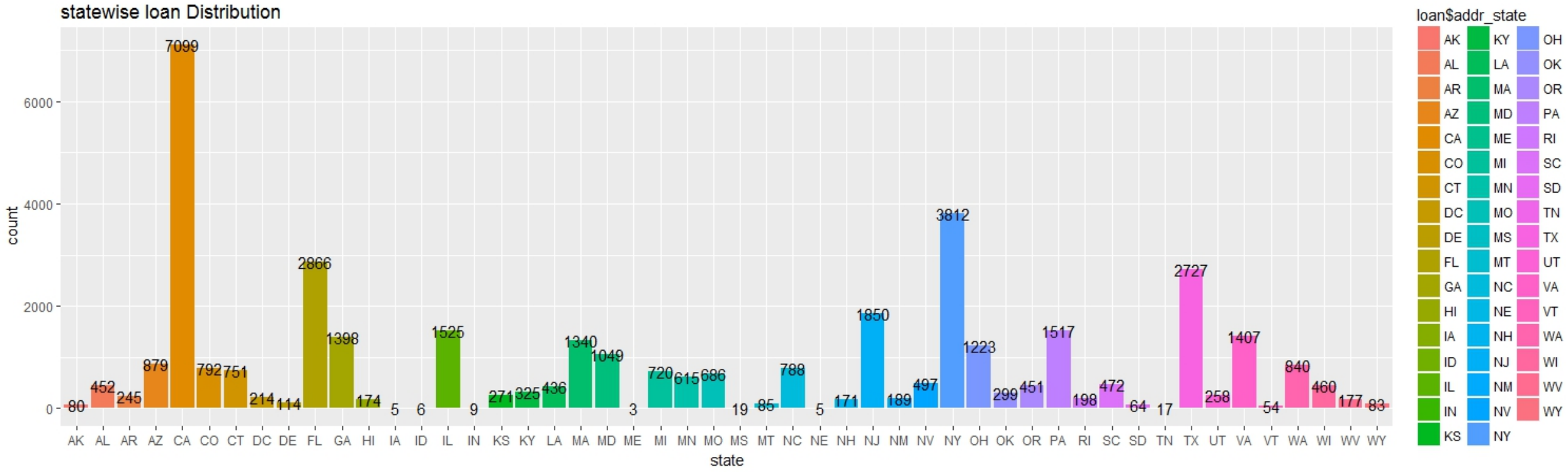




42.6% of Loans are not Verified

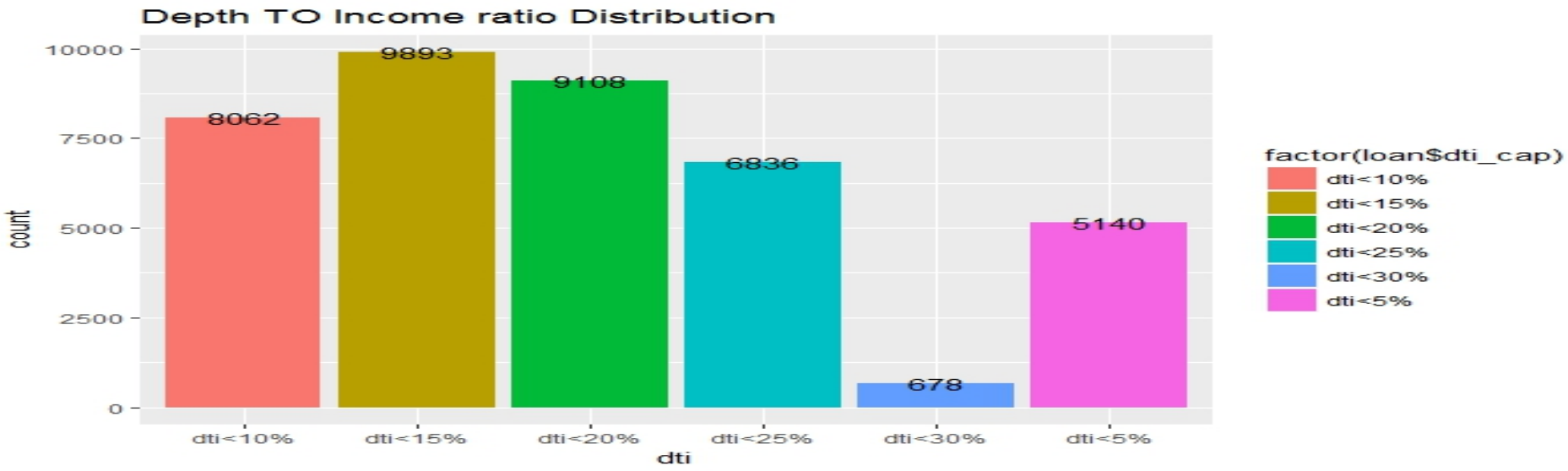
92% of People who have taken loan have Mortgaged or have a rented house

Univariate Analysis – State Analysis

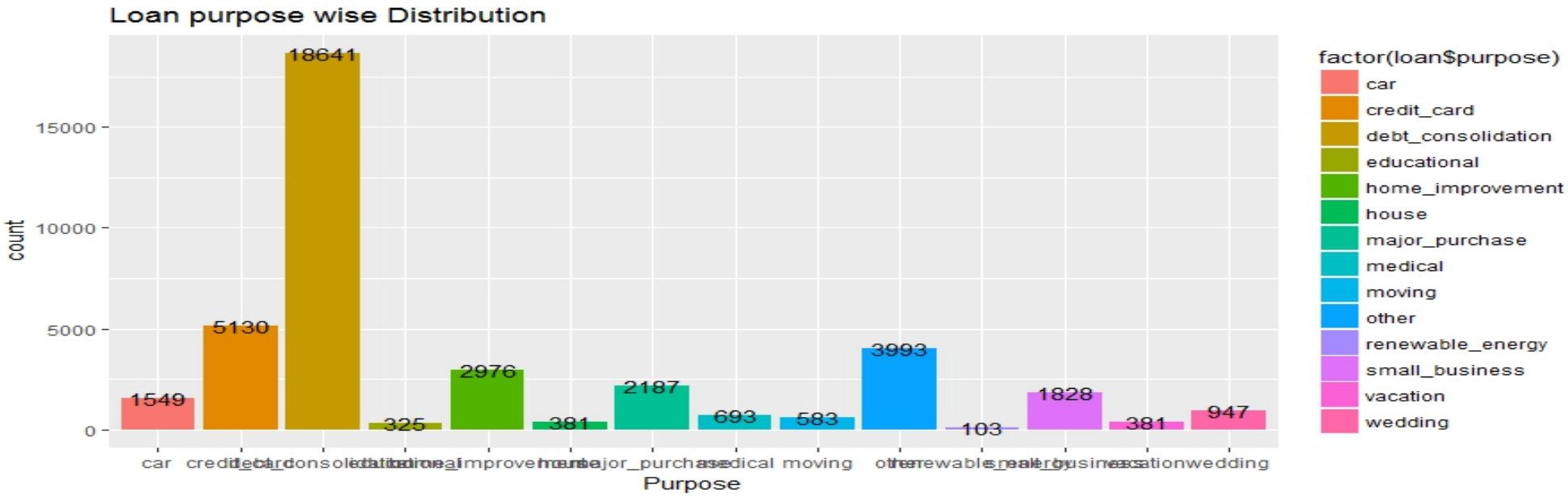


CA is the state which has seen the largest amount of loans i.e. 17.9 %

Univariate Analysis – Verification and Home ownership Analysis

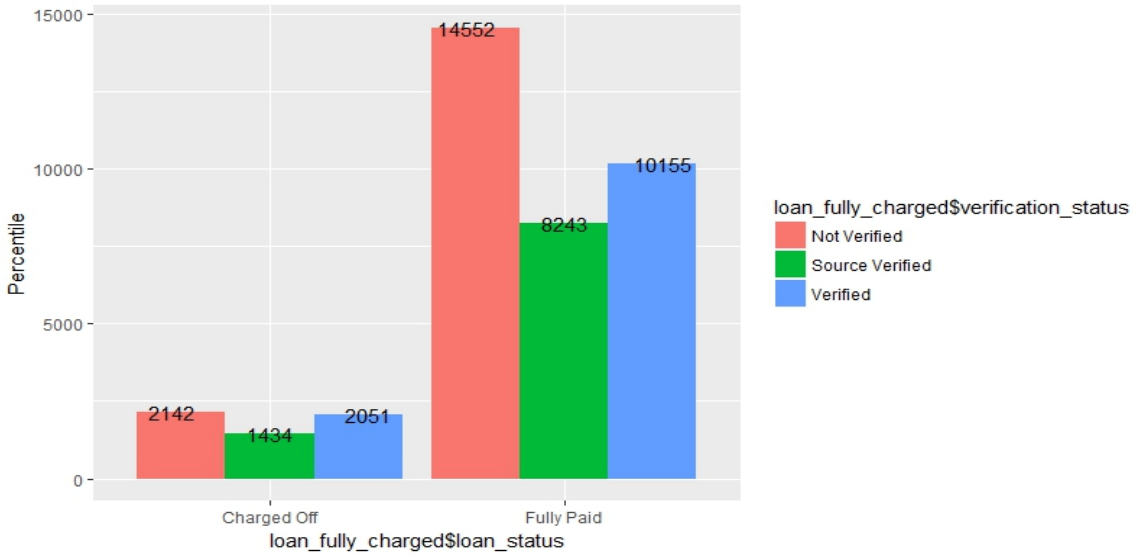


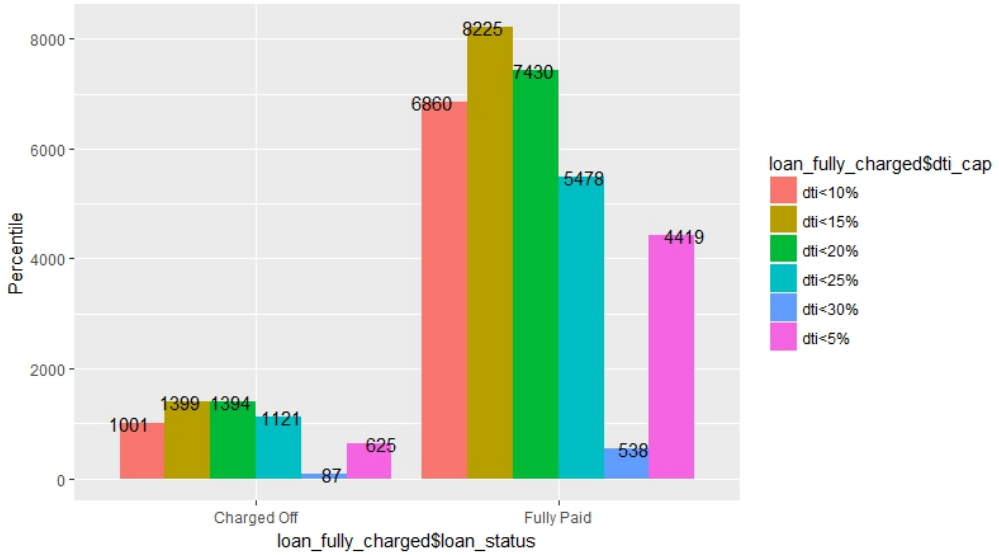
DTI distribution is Uniform for % between 0-30



46.7% of the loans are for Debt consolidation

Loan Status with Verification Status & DTI

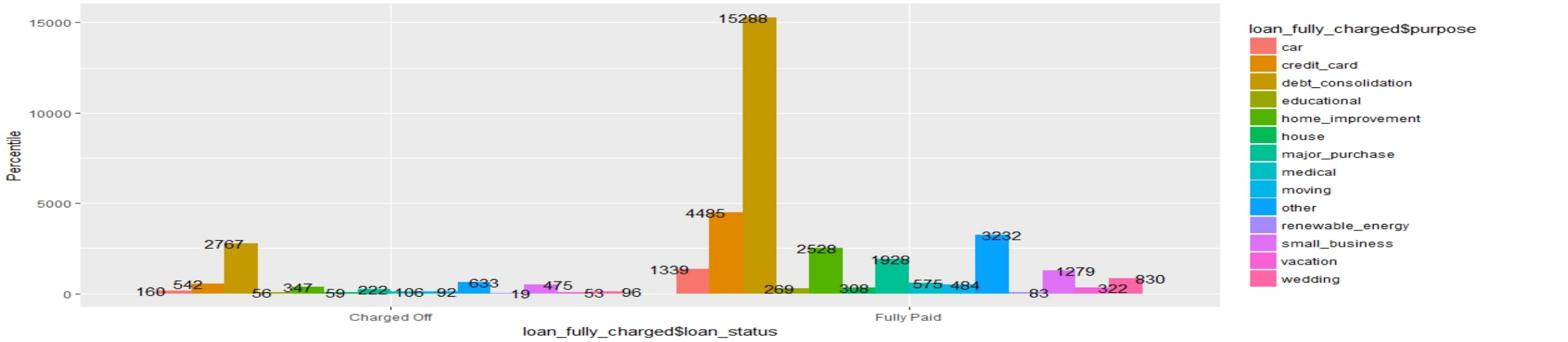


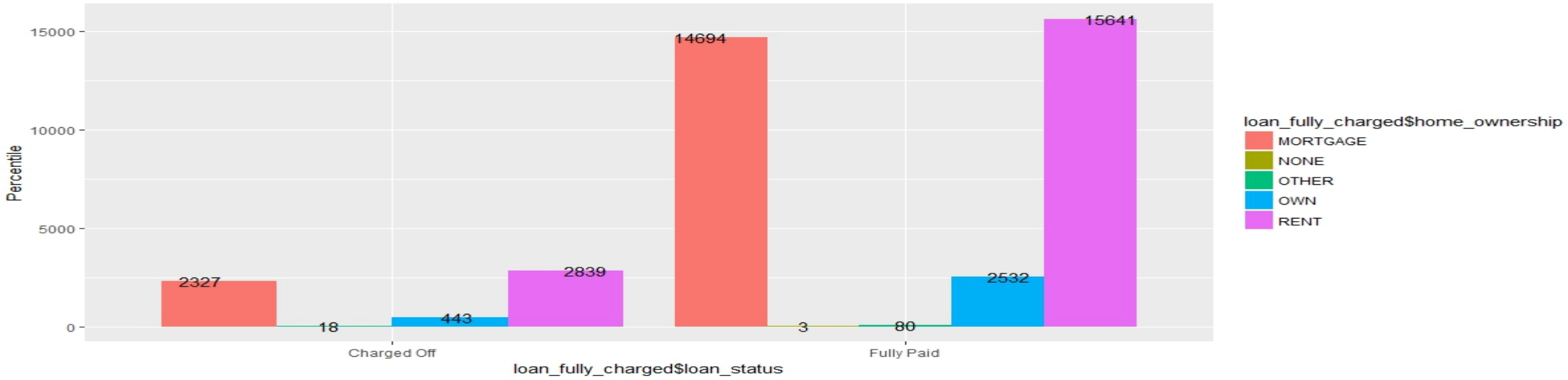


Verification Status shows similar trend
Between Fully Paid and Charged Off Loans

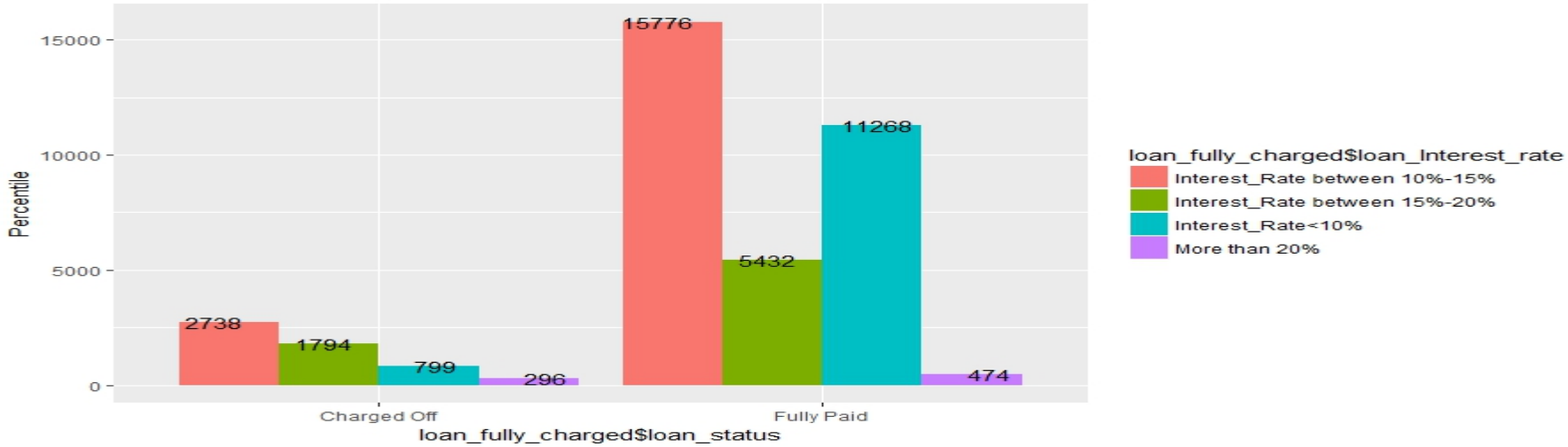
People with DTI between 10-25% have been charged Off

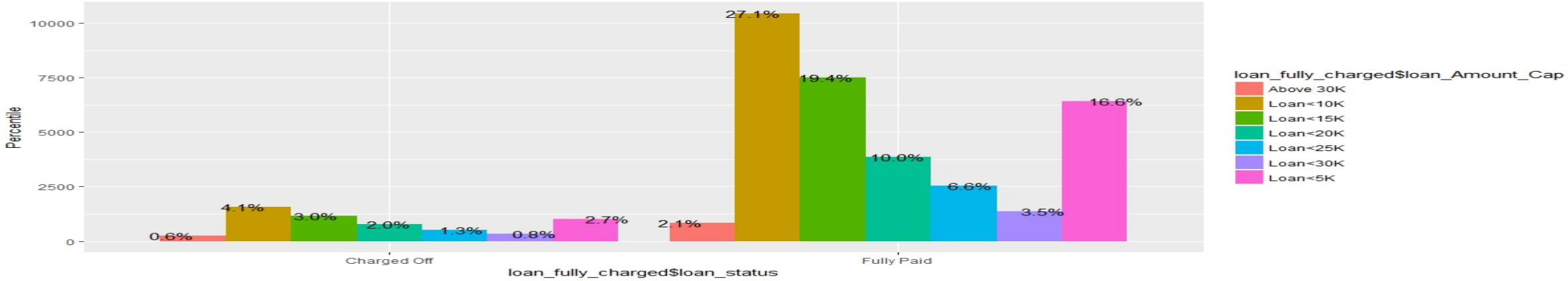
Loan Status with Purpose and Home OwnerShip



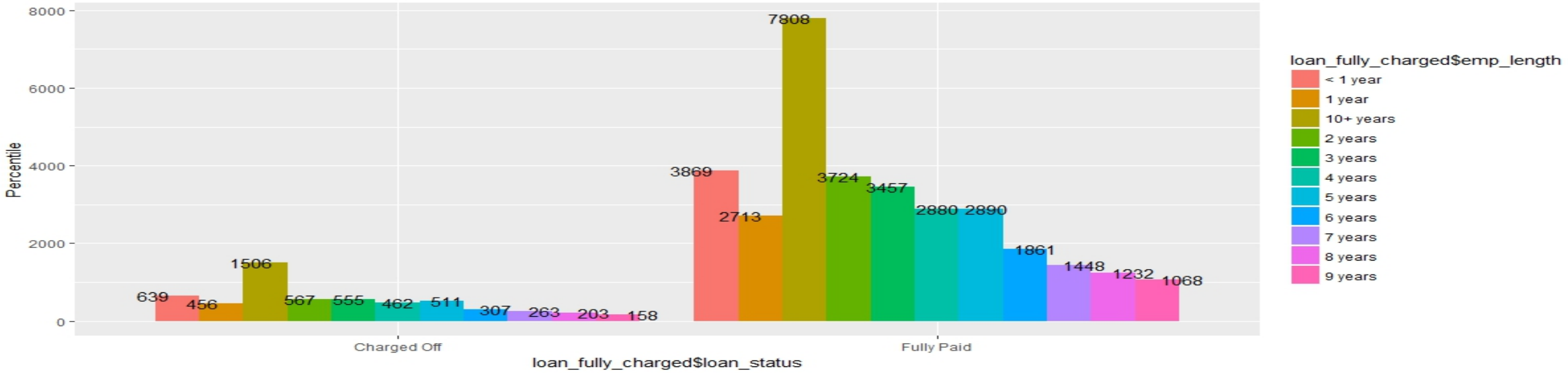


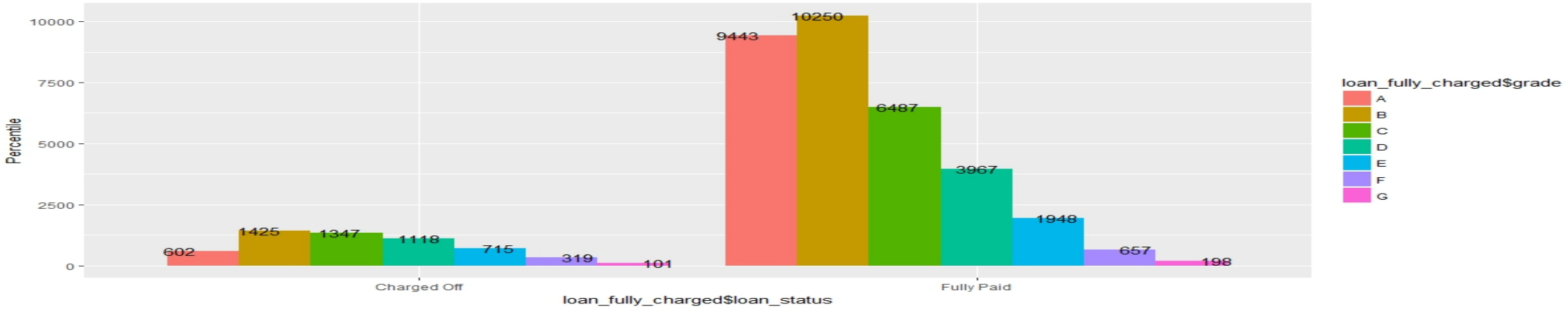
Loan Status with Loan Interest rate and Loan Amount



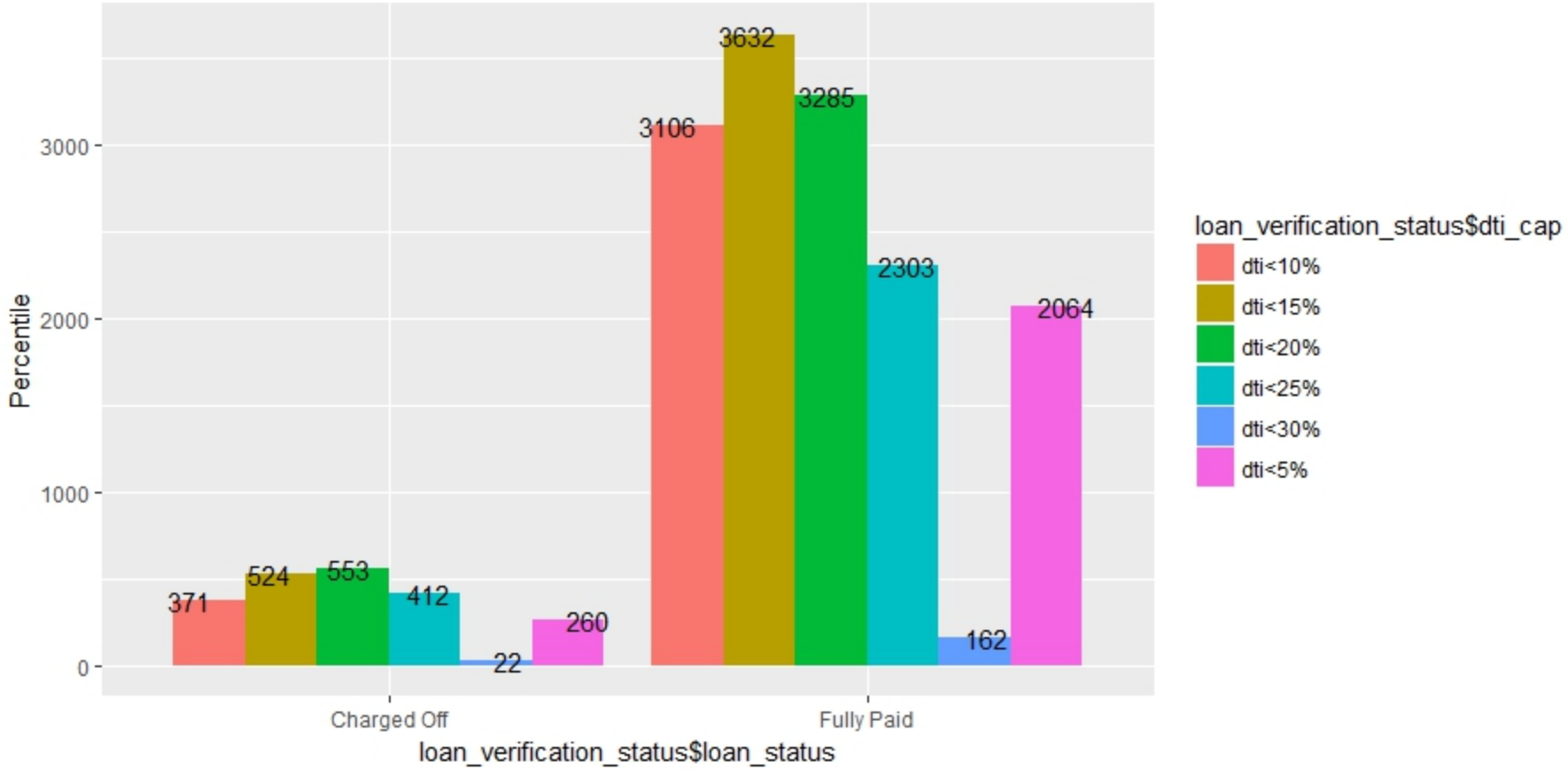


Loan Status with Employee Length and Grade

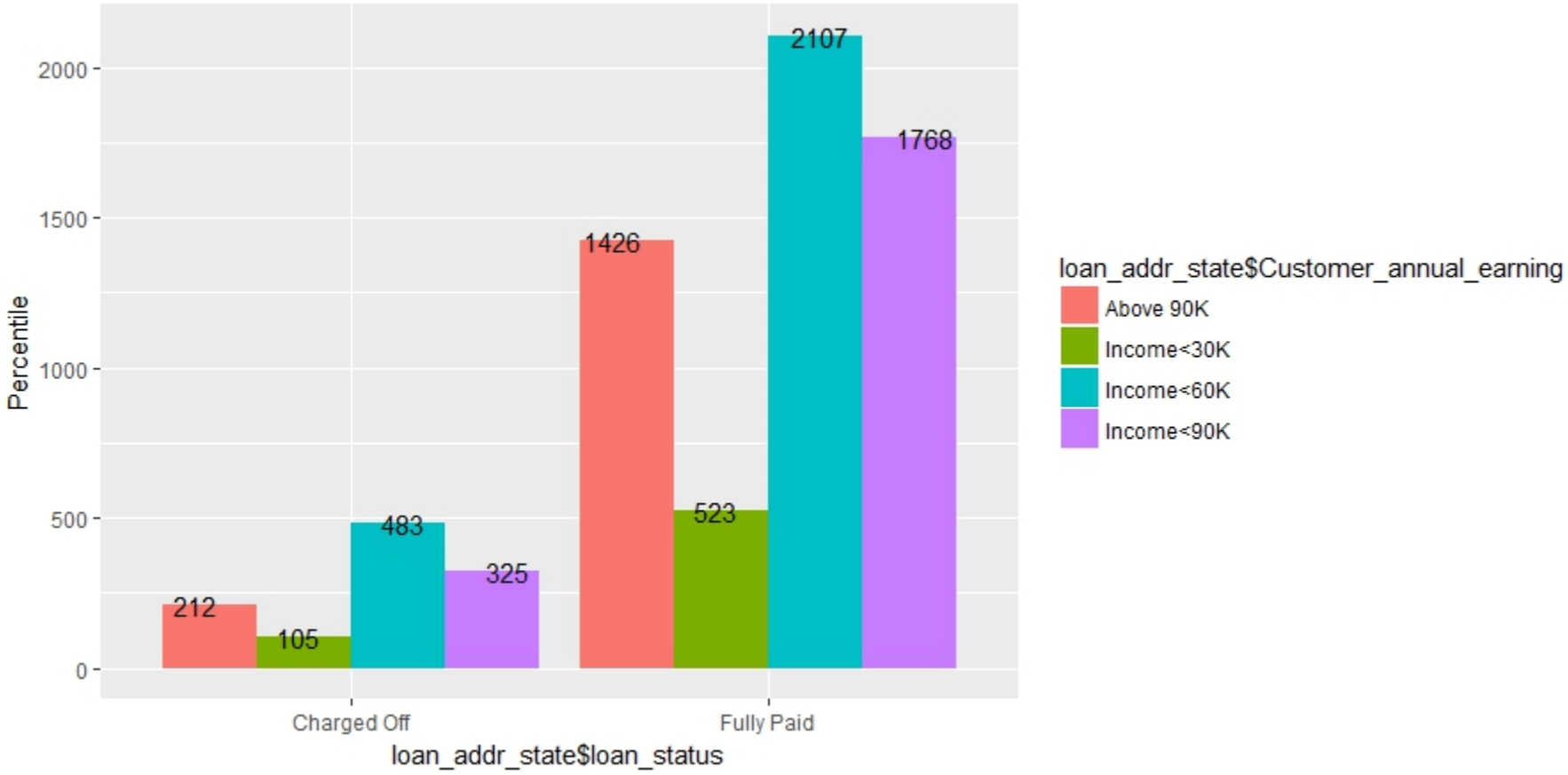




Multi Variate Analysis: loan status with For not Verified Source



Multi Variate Analysis : Loan Status for CA state



72 % of charged off in CA is between 30 k and 90k Annual Income

Conclusion

- 12% of the Non Verified Source have lead to Loan Charged off
- Around 50% of Loan Charged off is of type Debt_consolidation .
- Around 50% of people are given loan between 10-15%.
- Around 26% Charged Off are from emp_length 10+ years
- Grade B has seen the most amount of Charged off i.e. 1425