

# GRAMENER CASE STUDY

## SUBMISSION

**Group Name:**

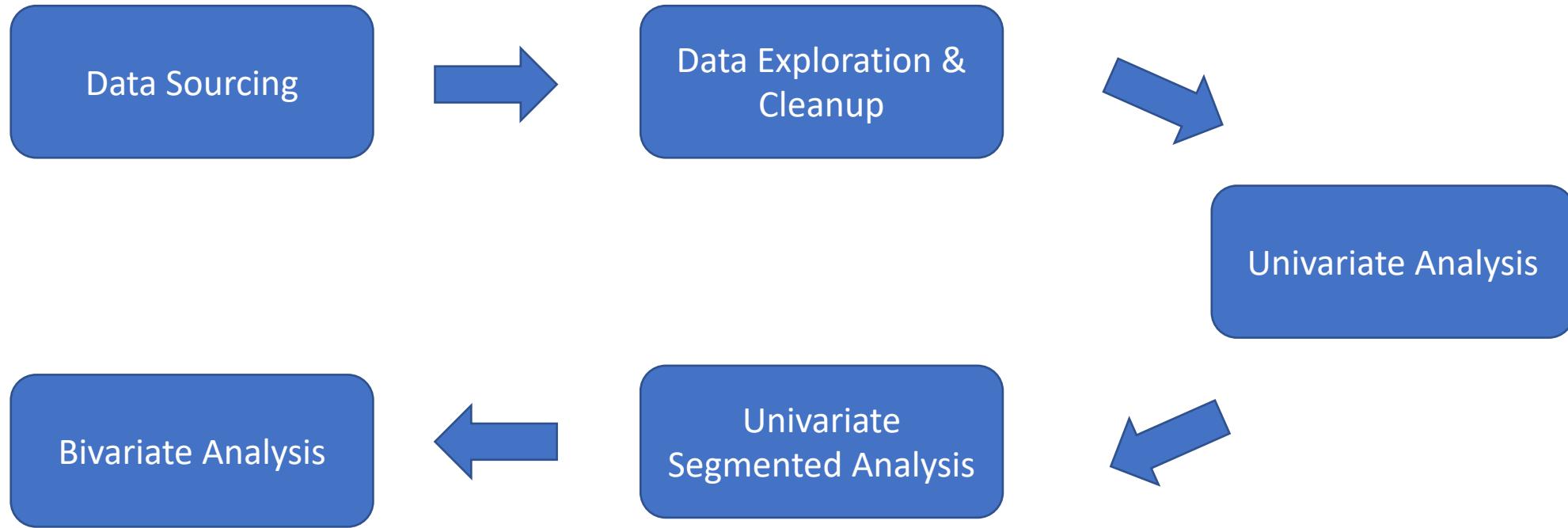
- 1. Mavallapalli Saraschandra**
- 2. Anuj Garg**
- 3. Varun Hasteer**
- 4. Prajwal Rao**

# Business Objective

- ❖ Identify “**High Risk**” loan applicants
- ❖ Understand how **consumer attributes** and **loan attributes** influence the tendency of default.
- ❖ Risk Analytics
  - Likely to repay the loan -> Results in a loss of business
  - Not Likely to repay the loan -> Lead to a financial loss
- ❖ Identify the patterns to make decisions
  - Reject Loans
  - Reduce the amount of loan
  - Lend at a higher interest rate



# EDA Workflow



# Data Sourcing

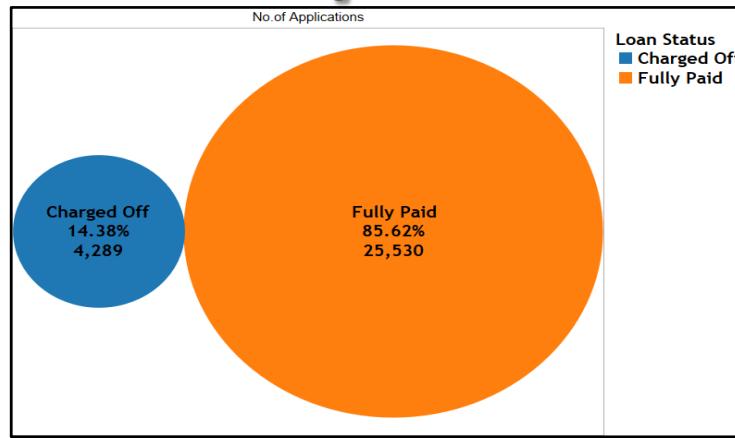
- ❖ Complete loan data for all loans issued through the time period 2007 to 2011
  - loan.zip
  
- ❖ Data dictionary which describes the meaning of variables
  - Data\_Dictionary.xlsx



# Data exploration

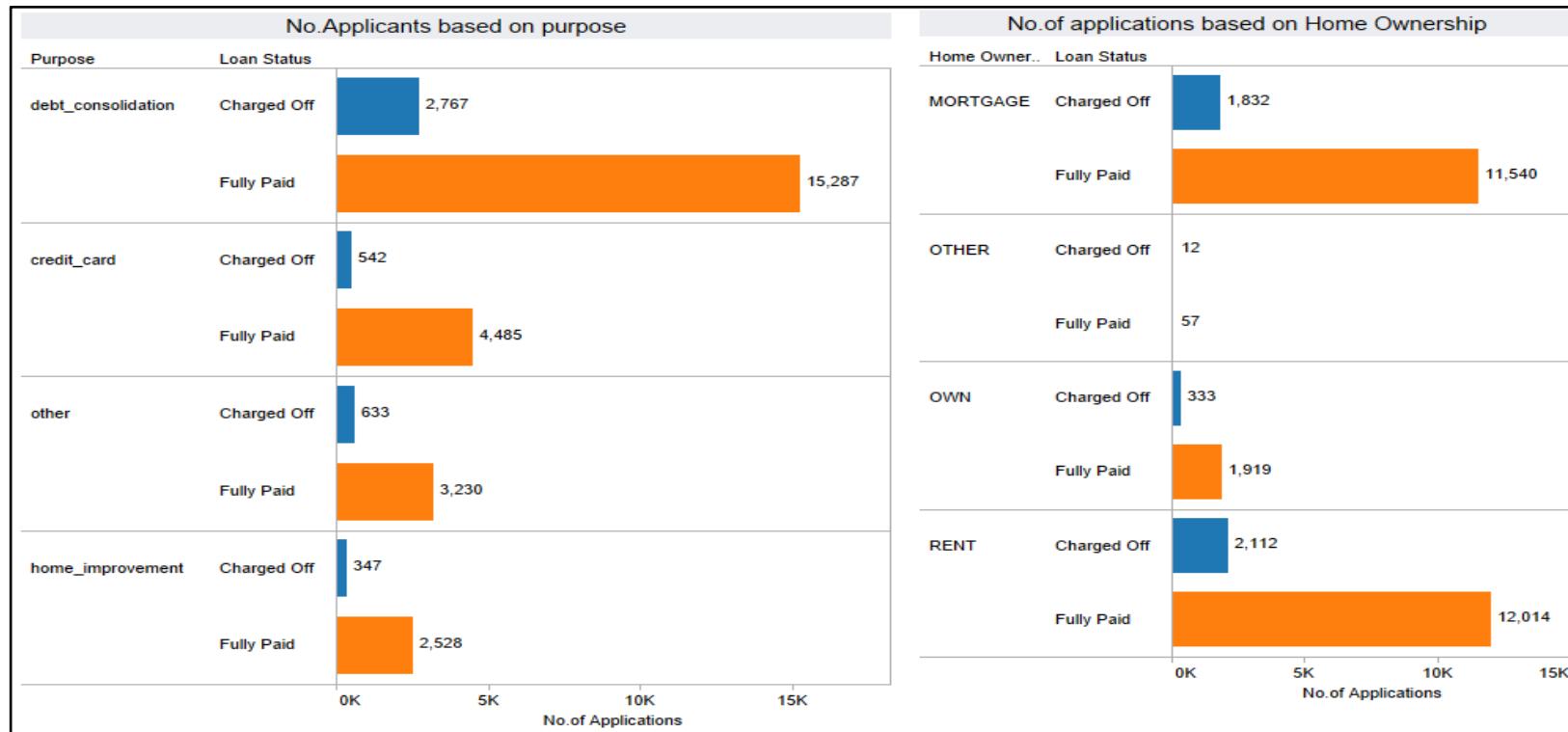
## ❖ Loan categories

- Charged Off ... 4,289 applicants
- Fully Paid. ... 25,530 applicants



## ❖ Insights

- Total Applicants ... 29819
- People paid their loans ... 85.62%
- Percent of defaulters ... 14.28%
- More Debtors are defaulted with Debt consolidation Purpose Loan
- More Debtors are defaulted having Rented Home or Mortgage Home



# Data Clean-up

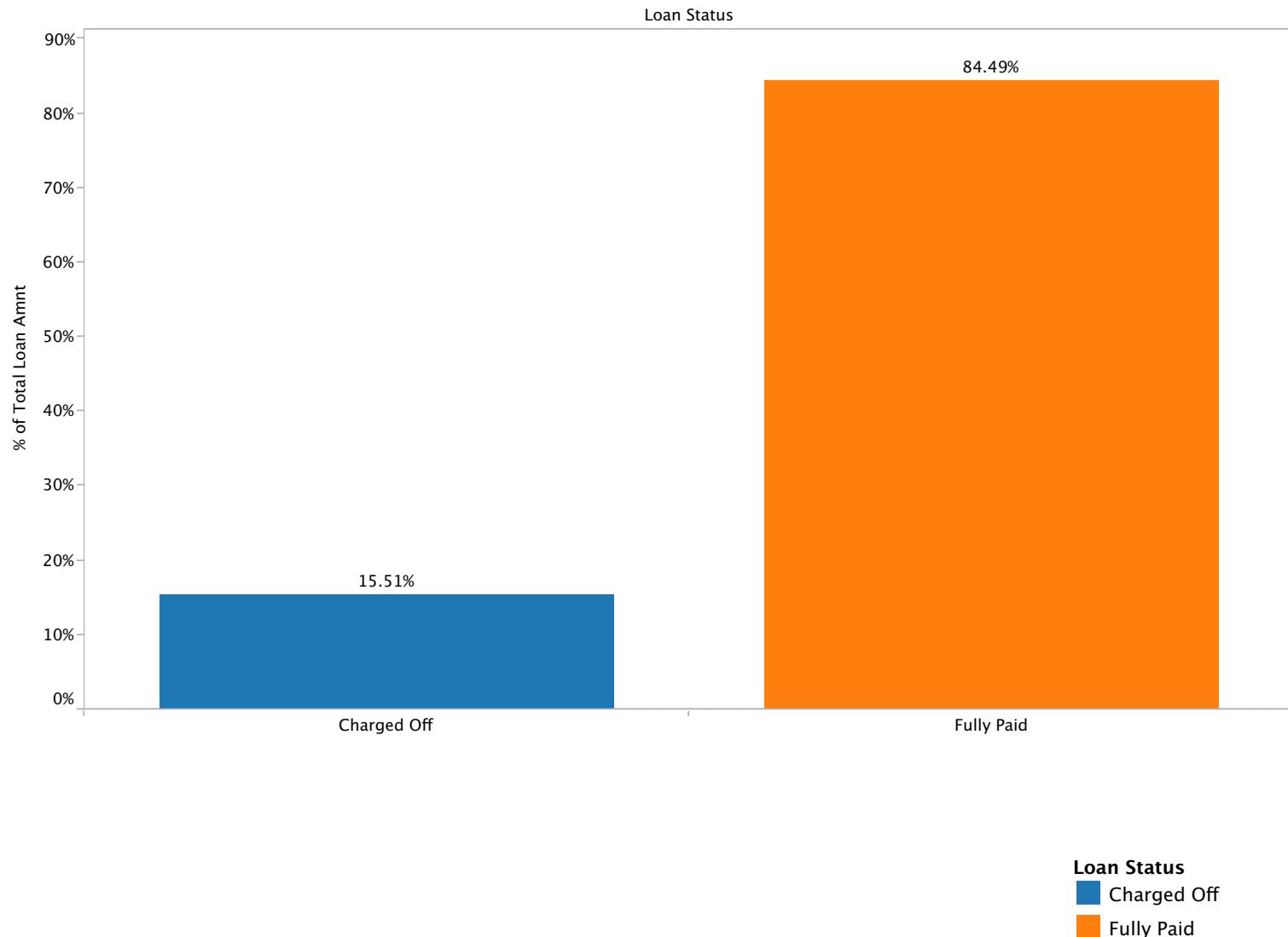
Task	Value	Applied action
Total (Rows, Columns) identified in the dataset	39717, 111	Imported columns for data analysis
Maximum 100 % of missing values in a column	60	Removed columns from the analysis
Columns which have 0 values	4	Removed columns from the analysis
Columns which are not required for analysis	24	Removed columns from the analysis
Special character symbol in the int_rate column	%	Removed % from the column
String to Date Format conversion	%B-%Y	Converted String to Date Format (%B-%Y)
Identify np.NAN in emp_length column	NAN	Substituted by 0
Conversion of emp_length column from string to int	+, <	Removed + and < character, convert to numeric
Rows with "Current" loan status	loan_status = Current	Removed rows form the analysis
Total (Rows, Columns) for data analysis	38577, 23	Available for data analysis

# Univariate Analysis

## ❖ Loan Status vs Total Loan Amount

Fully Paid	84.49 %
Charged Off	15.51 %

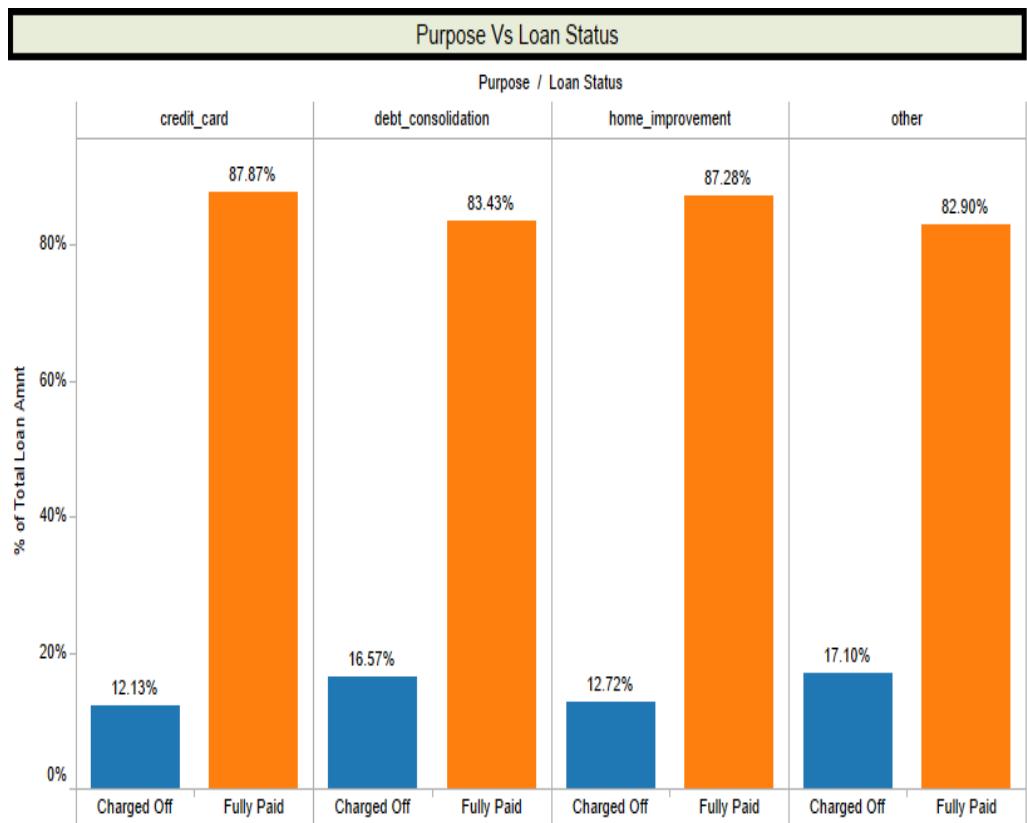
Defaulters are 15.51%.



## ❖ Purpose vs Loan Status

**Others : 17.10 % defaulters**

**Second largest in Debt Consolidation : 16.57  
defaulters**

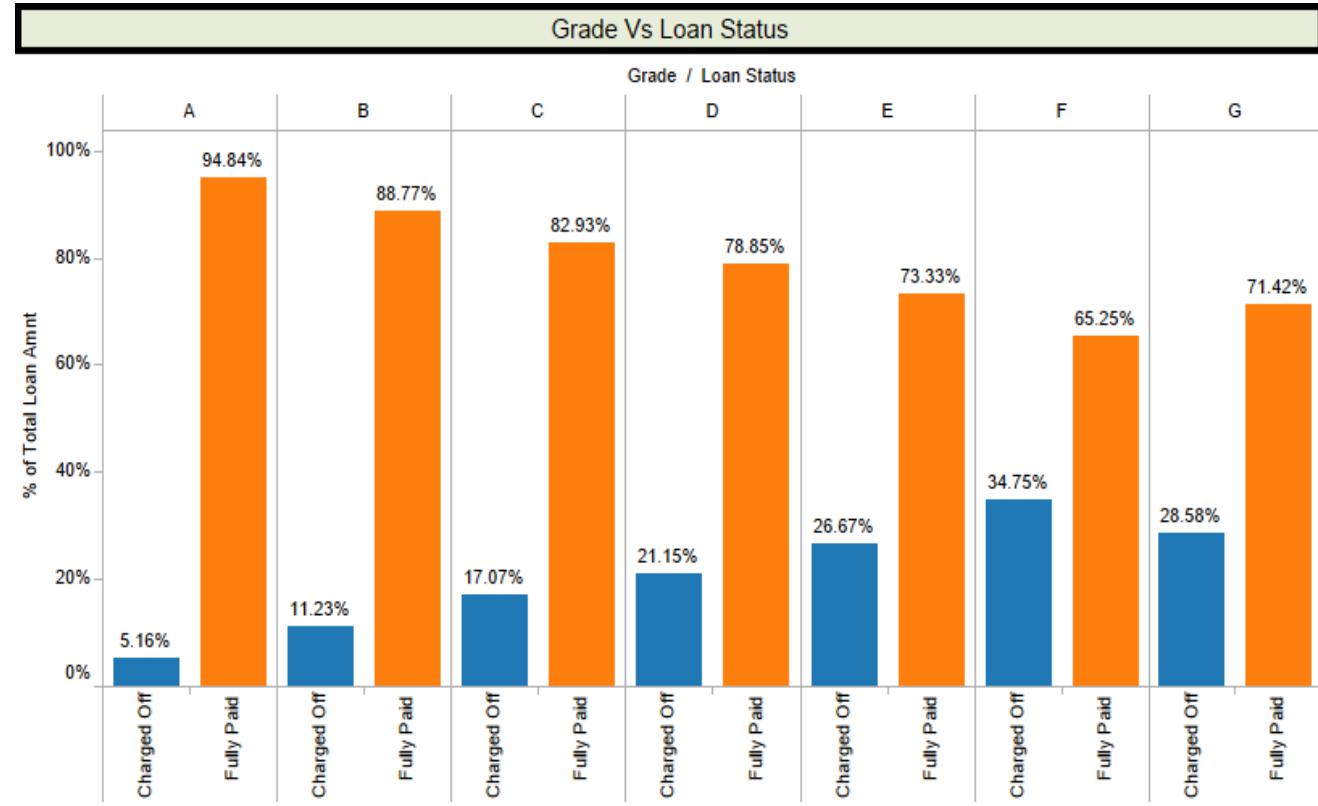


## ❖ Grade vs Loan Status

**F Grade** loans are the most defaulters : **34.75 %**

**A Grade** loans are the most paid : 94.84 %

A lower grade shows very high correlation with the applicant being a defaulter.

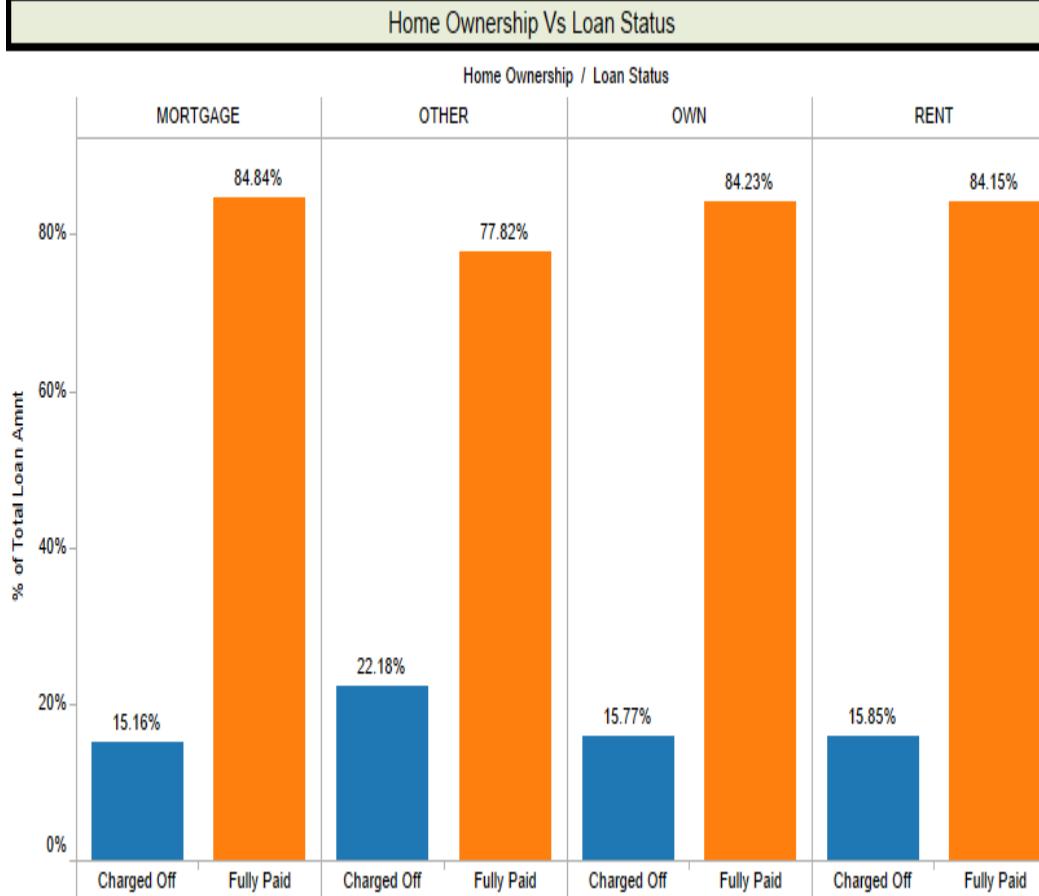


# Segmented Univariate Analysis

## ❖ House Ownership vs Loan Status

Others : 22.18 % defaulters

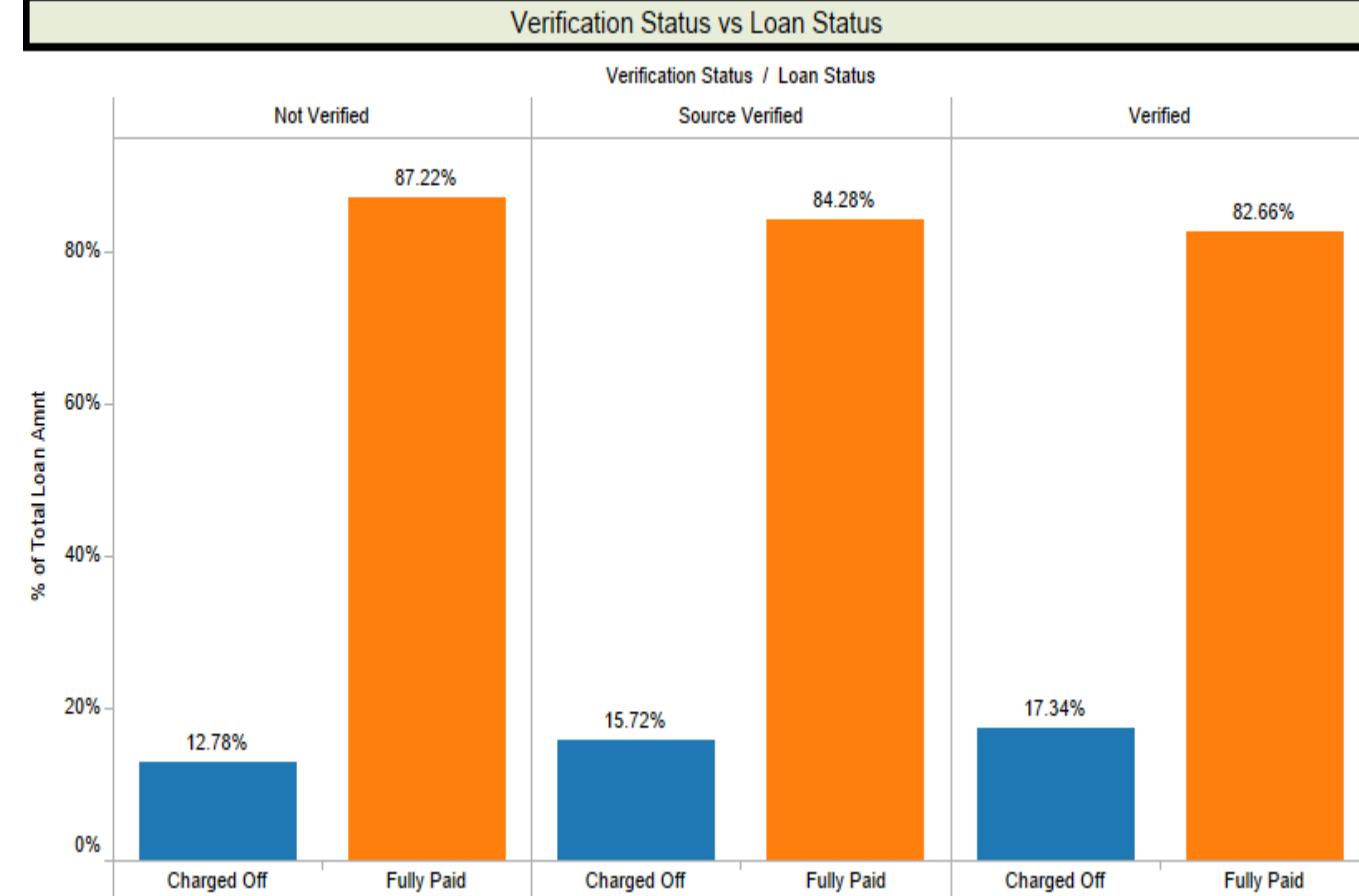
Home Ownership Vs Loan Status



## ❖ Verification Status vs Loan Status

Verified loans are more defaulters : 17.34 % defaulters

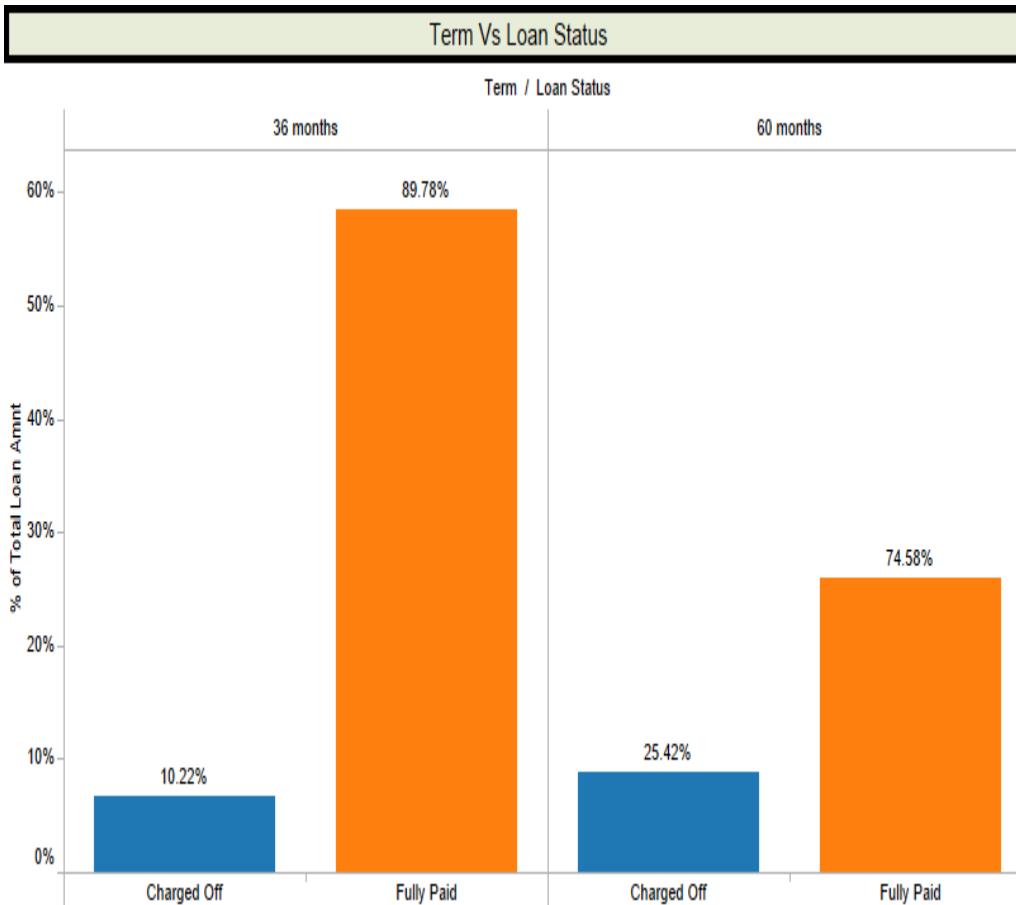
Verification Status vs Loan Status



# Segmented Univariate Analysis

## ❖ Term vs Loan Status

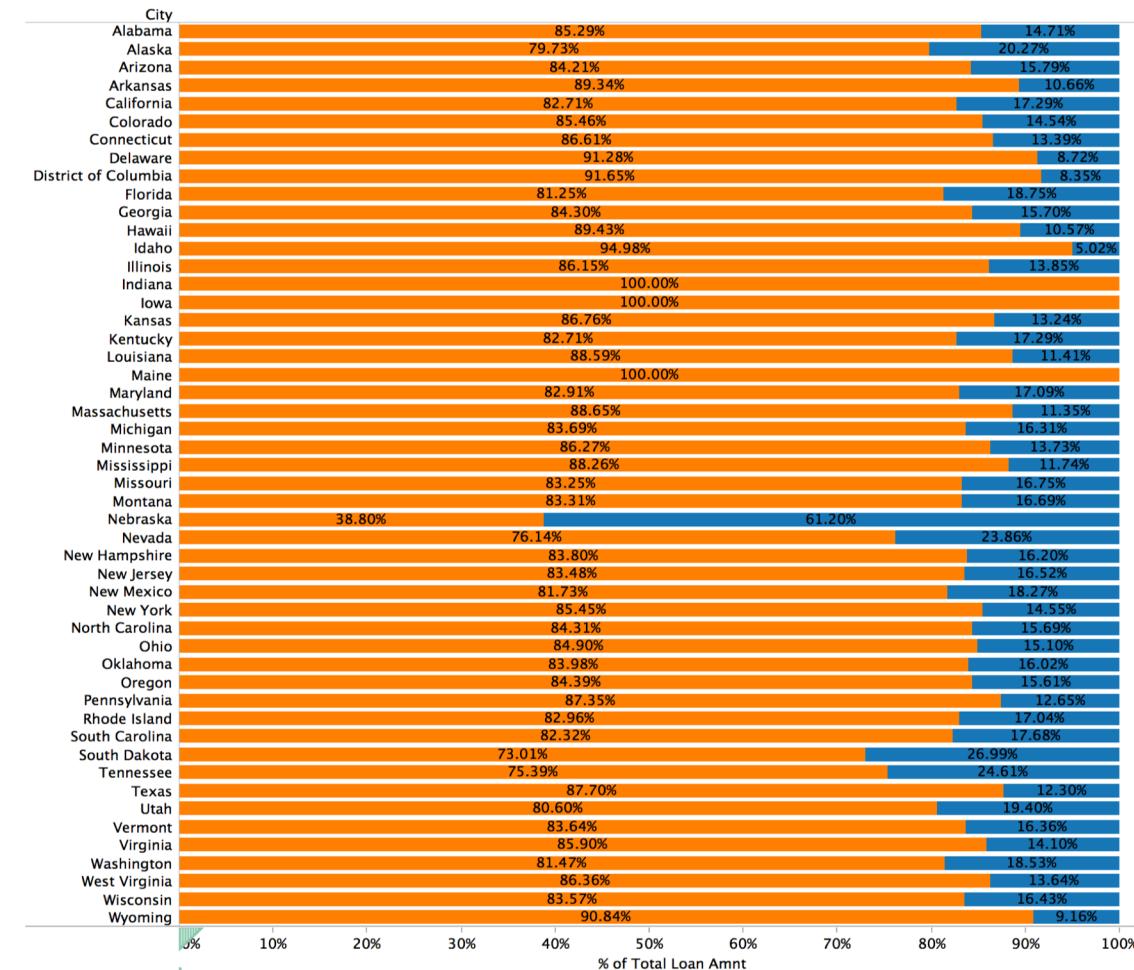
More defaulters in 36 month loan : **58.53 %**



Largest defaulters (61.20 %) are in **"Nebraska"** state. It is outlier. Hence, it is not considered.

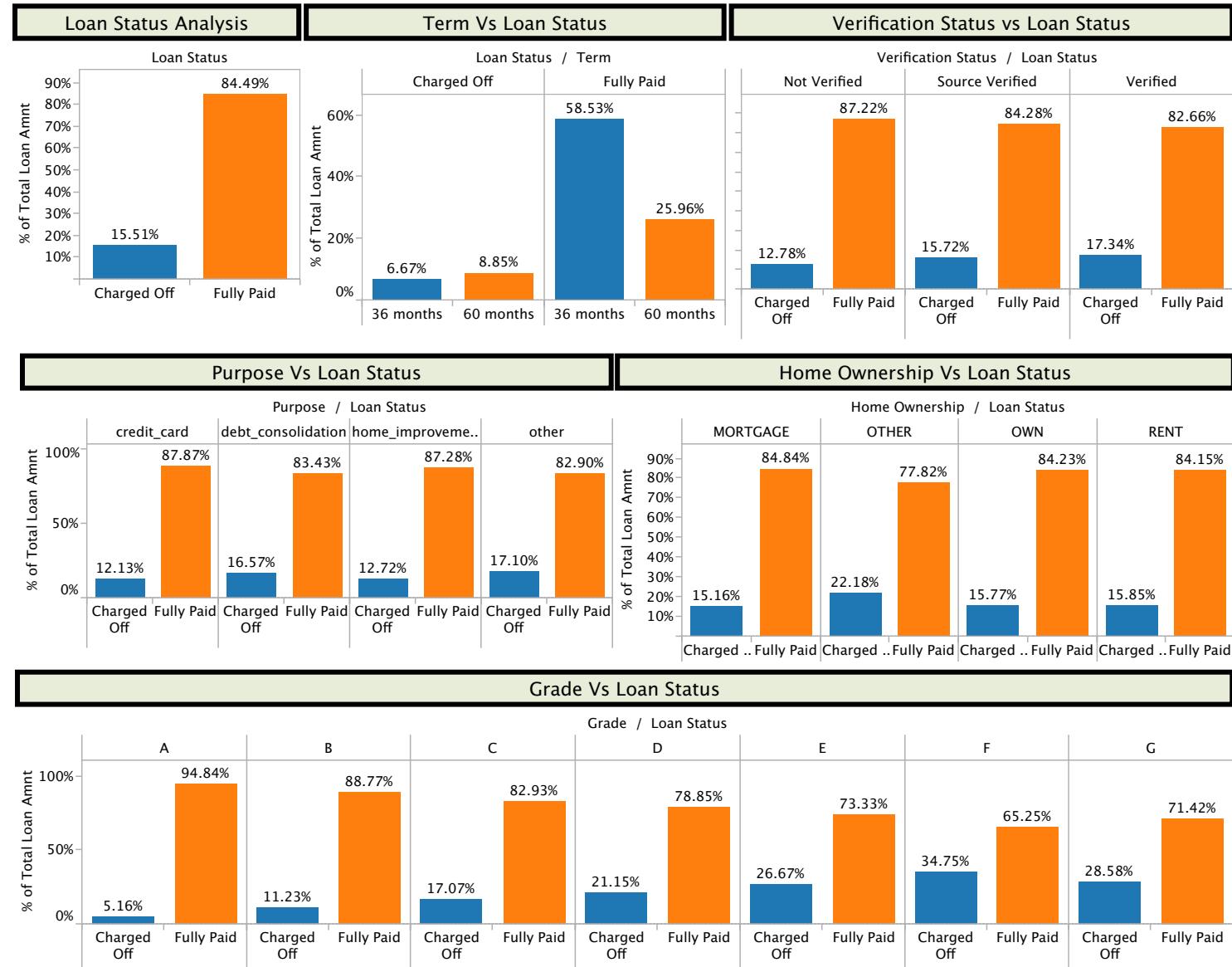
Second largest defaulters (26.99 %) are in **"South Dakota"**

## ❖ Verification Status vs Loan Status



# Segmented Univariate Analysis

1. Term vs Loan Status shows the strong relationship.
2. Verification vs Loan Status shows very strong relationship. Maximum defaulters are with verified status.
3. Purpose vs Loan Status shows good relationship because maximum defaulters are for debt consolidation.
4. Grade vs Loan Status shows the relationship where F grade has maximum defaulters.
5. Home ownership vs Loan Status shows borrowers market value.

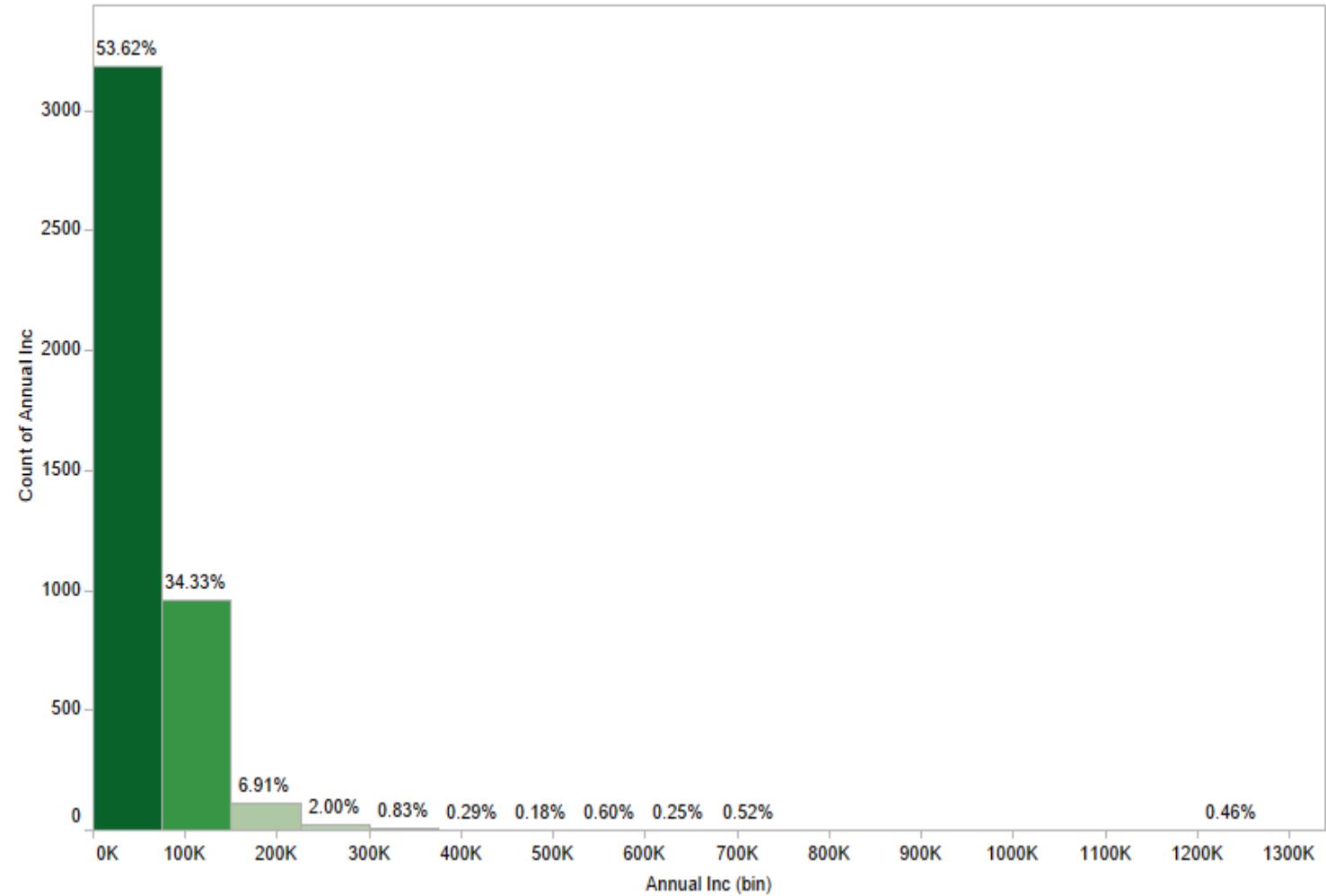


# Univariate Analysis

## ❖ Annual Income Analysis

Most of defaulters having annual income less than \$200K

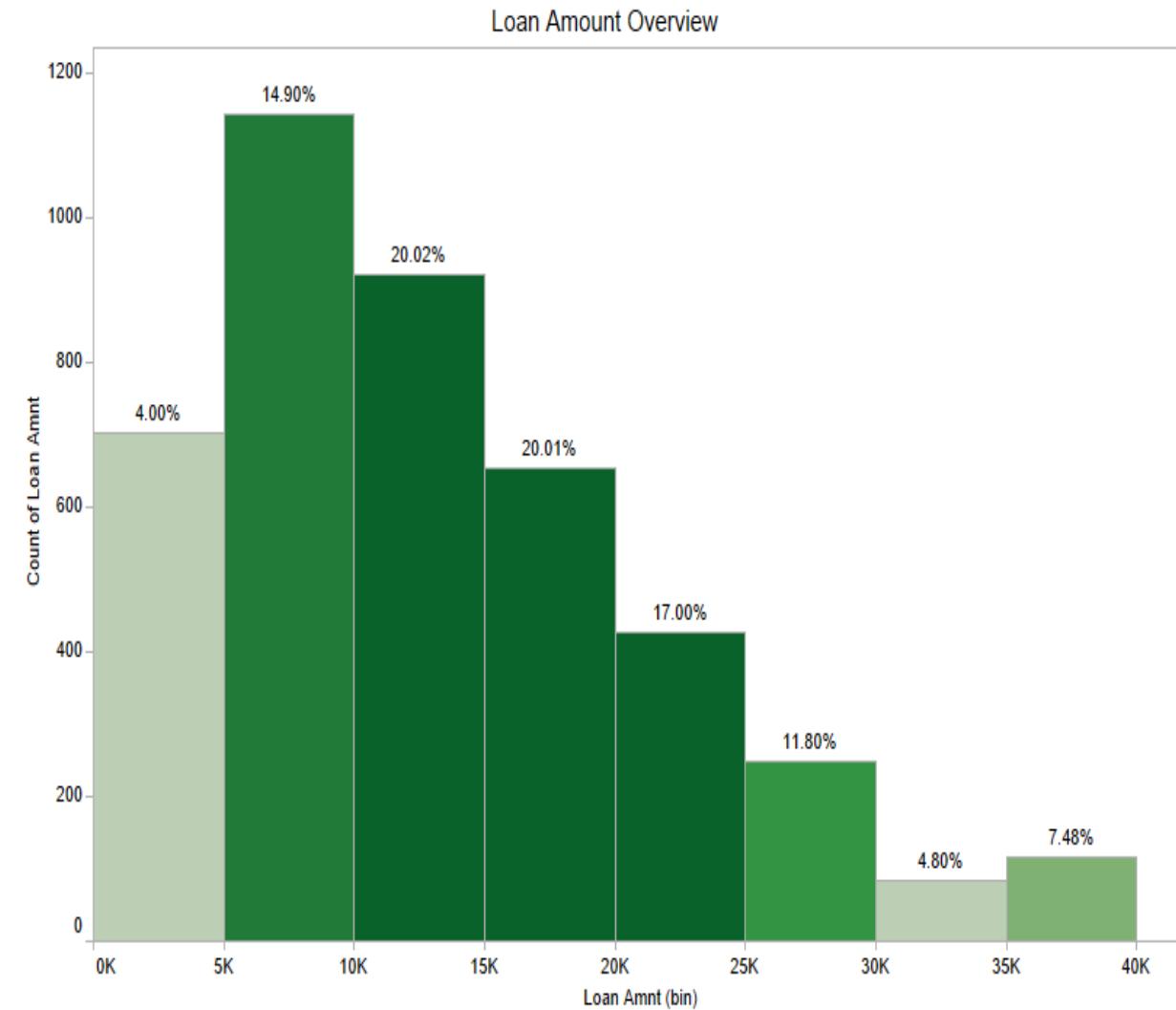
Annual Income Overview



# Univariate Analysis

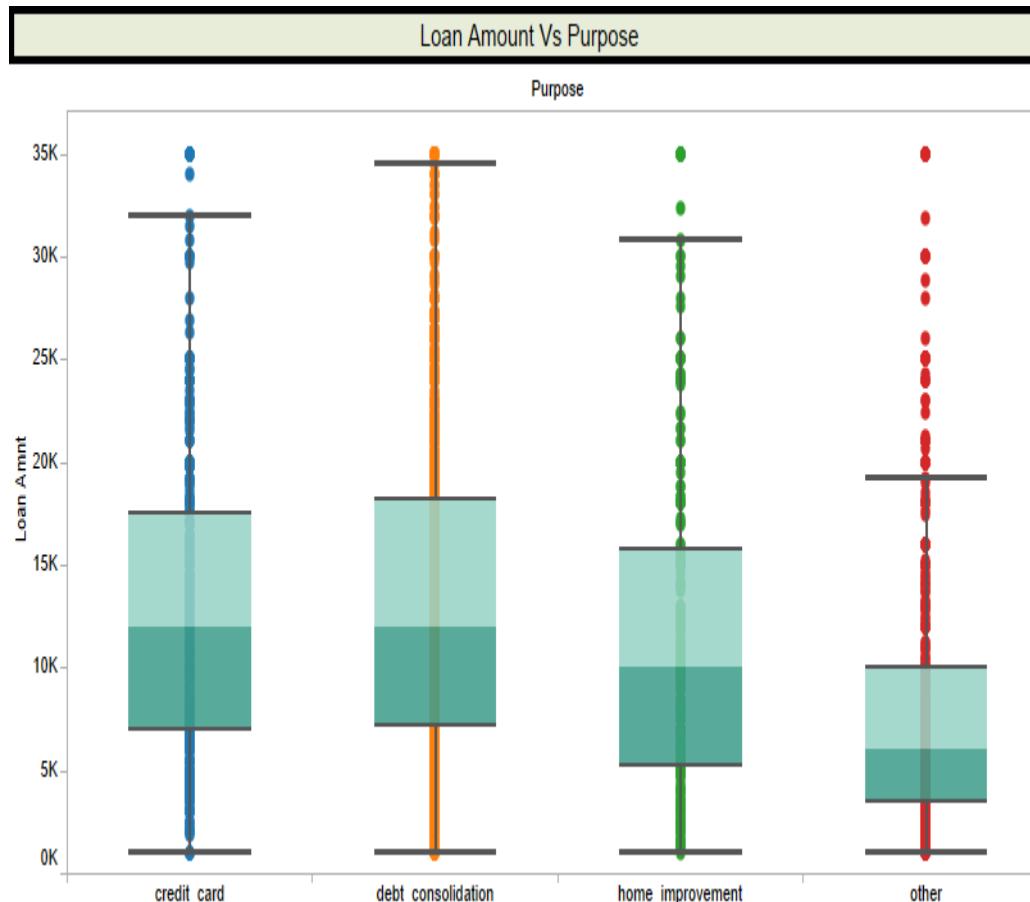
## ❖ Loan Status vs Total Loan Amount

Most of the applicants get Loan in between **\$5K to \$15K**



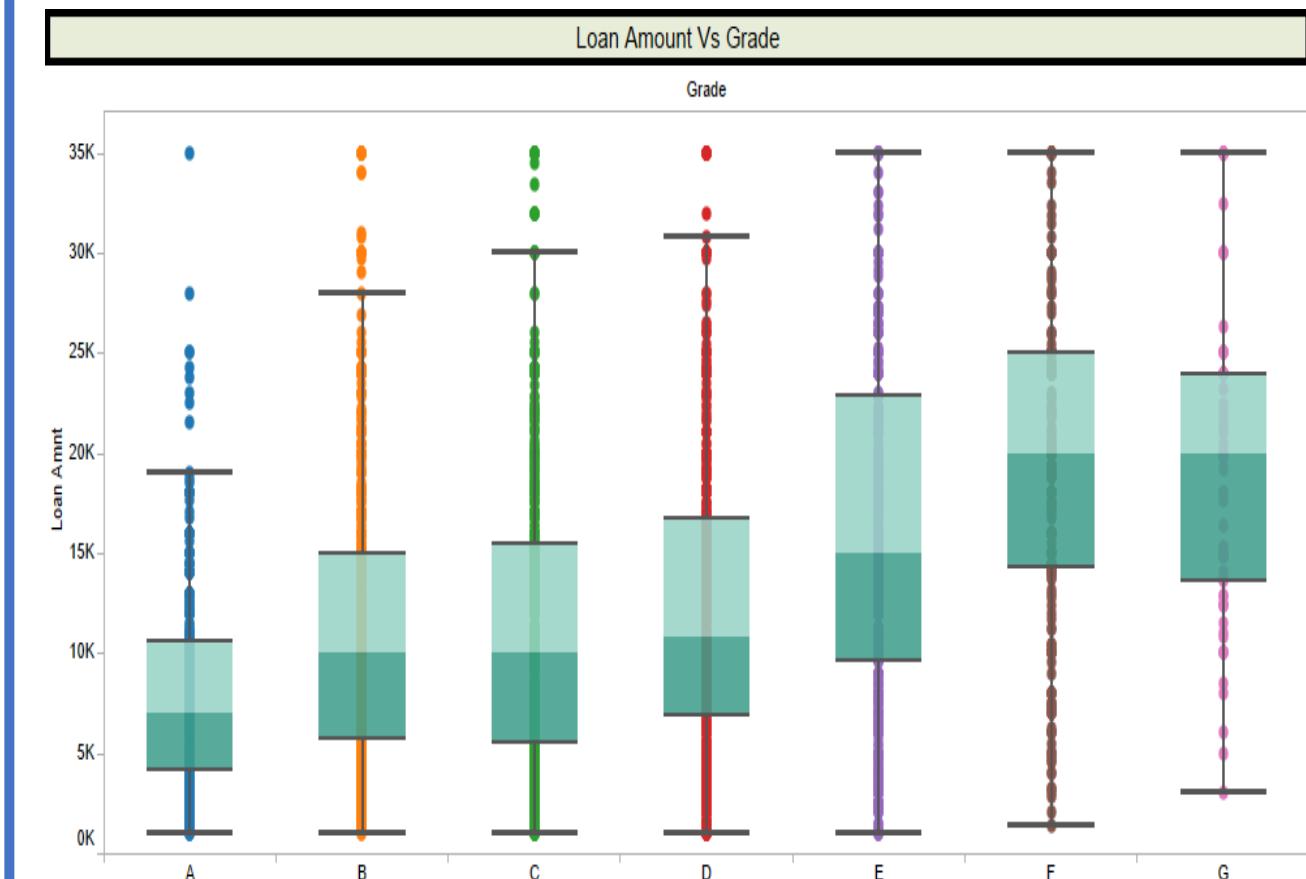
# Bivariate Analysis

## ❖ Loan amount vs Purpose



It has been observed that major loan amount defaulted in case of other purpose loan.

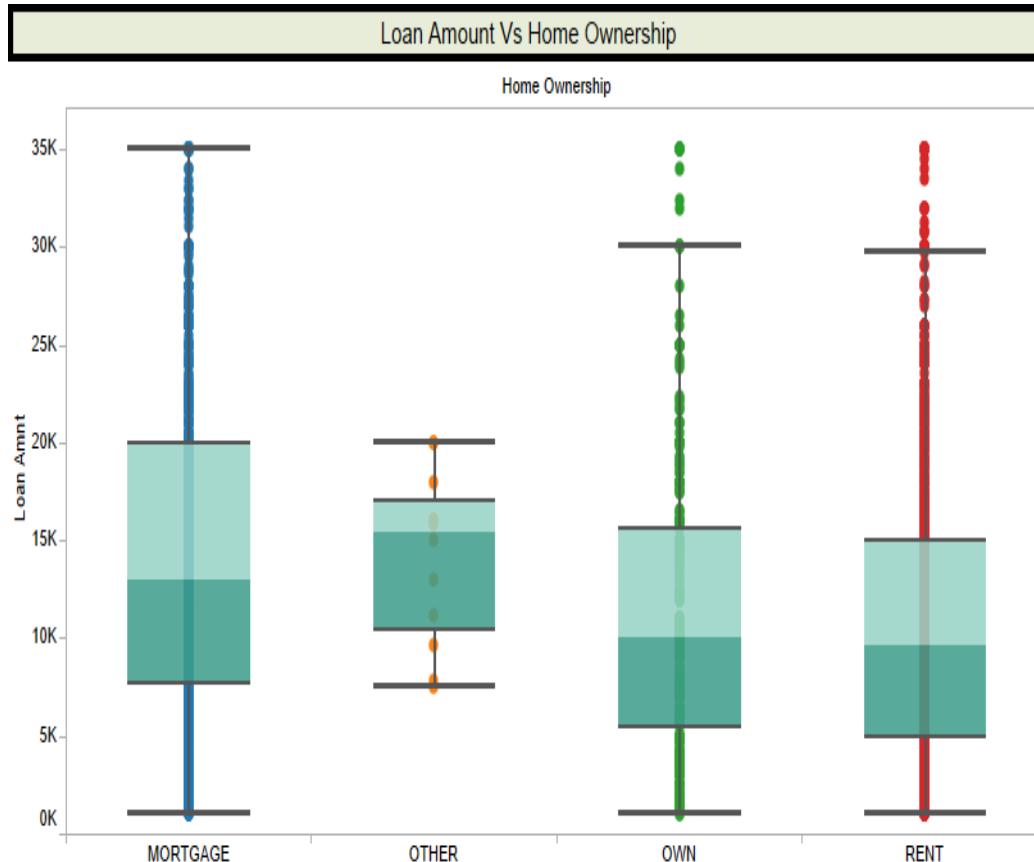
## ❖ Loan amount vs Grade



Strong correlation between Loan amount and Grade.

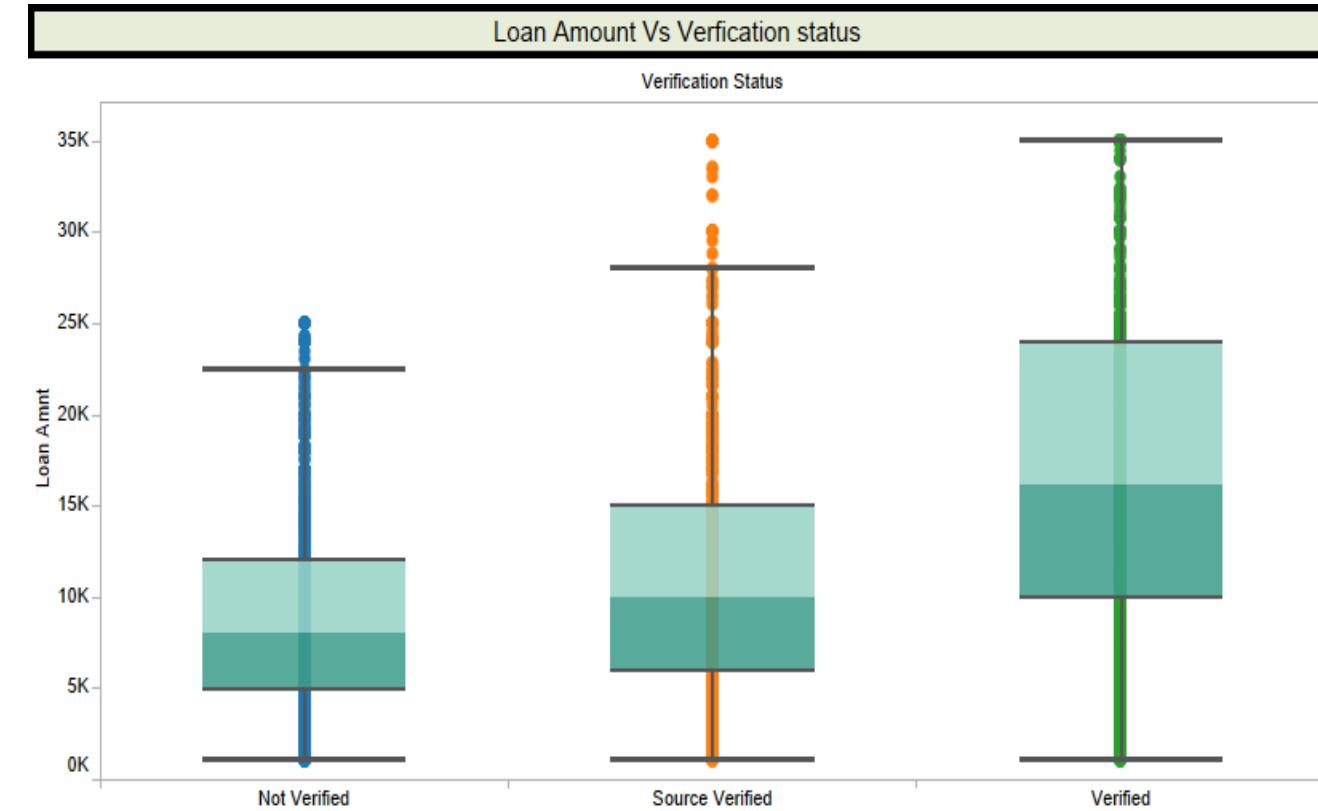
# Bivariate Analysis

## ❖ Loan amount vs Home Ownership



It has been observed that major loan amount is sanctioned for those who have rented house

## ❖ Loan amount vs Verification Status

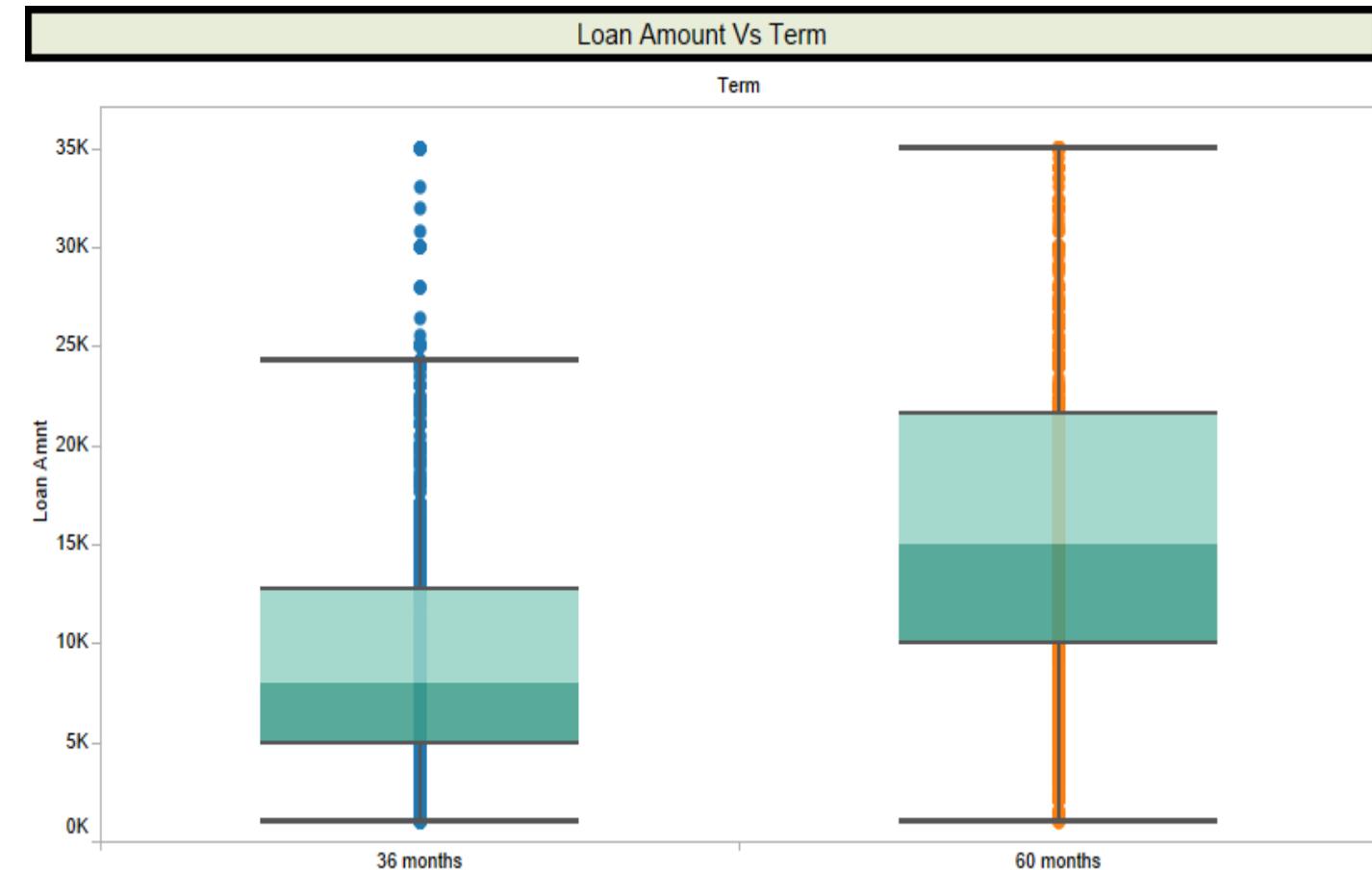


Verified loan status has the highest major loan amount

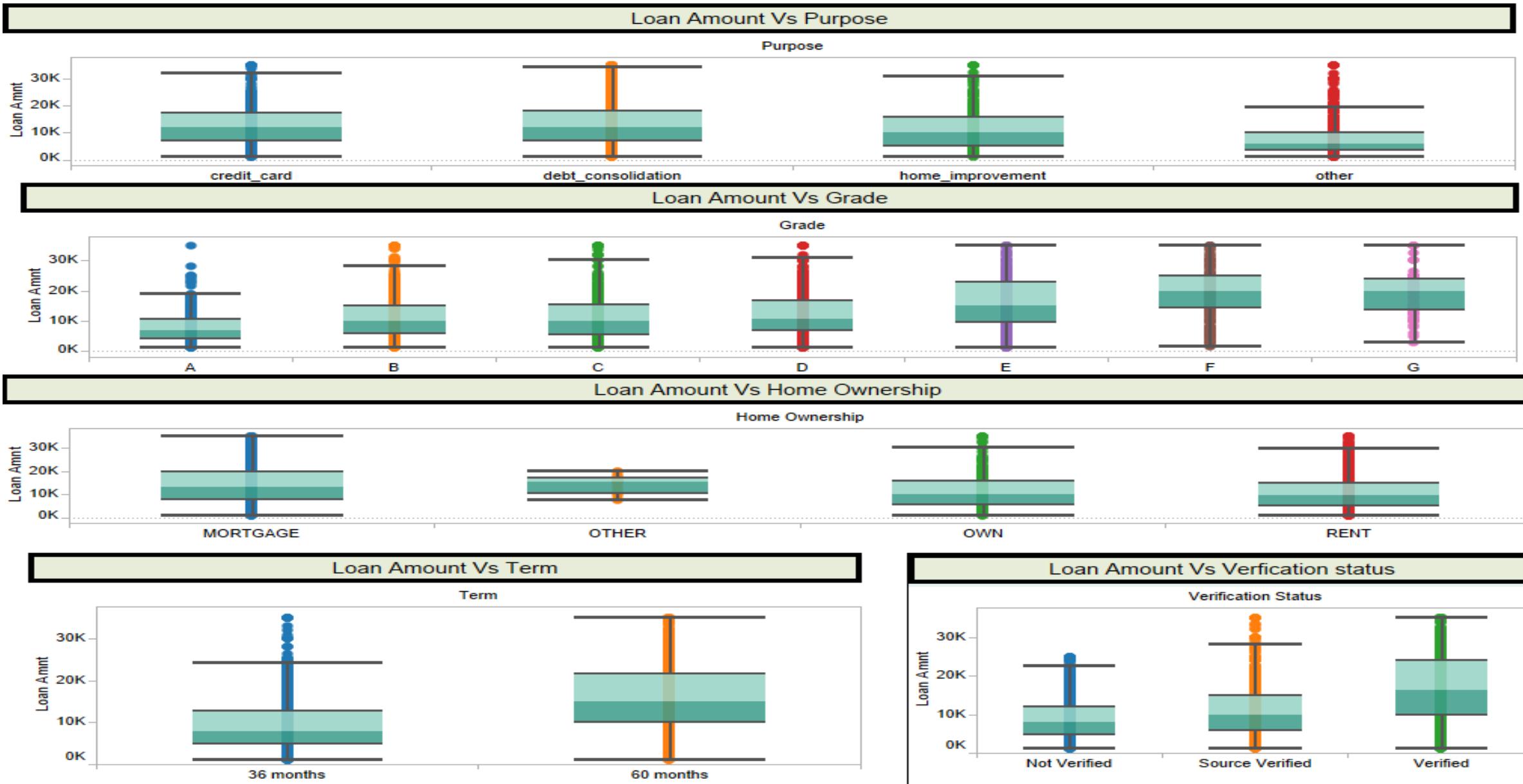
# Bivariate Analysis

## ❖ Loan amount vs Term

Major loan amount defaulted has been observed In 36 months term



# Bivariate Analysis



## ❖ Conclusion

- It has been observed that major loan amount defaulted in case of other purpose loan.
- More loan amount is observed in grade A.
- It has been observed that major loan amount is sanctioned for those who have rented house
- Verified loan status has the highest major loan amount
- Major loan amount defaulted has been observed In 36 months term

# Conclusion

- ❖ Borrower's with following factors are highly prone to be defaulters
  - Home Ownership
  - Mortgage
  - Rent
- ❖ Loan Purpose
  - DEBT\_CONsolidation
  - CREDIT\_CARD
  - HOME\_IMPROVEMENT
  - SMALL\_BUSINESS
  - OTHER
- ❖ Annual Income falling in range from **\$20,000 to \$1,00,000**
- ❖ Higher Debt to Income(DTI) ratio