

# THE BUSINESS PROFILE

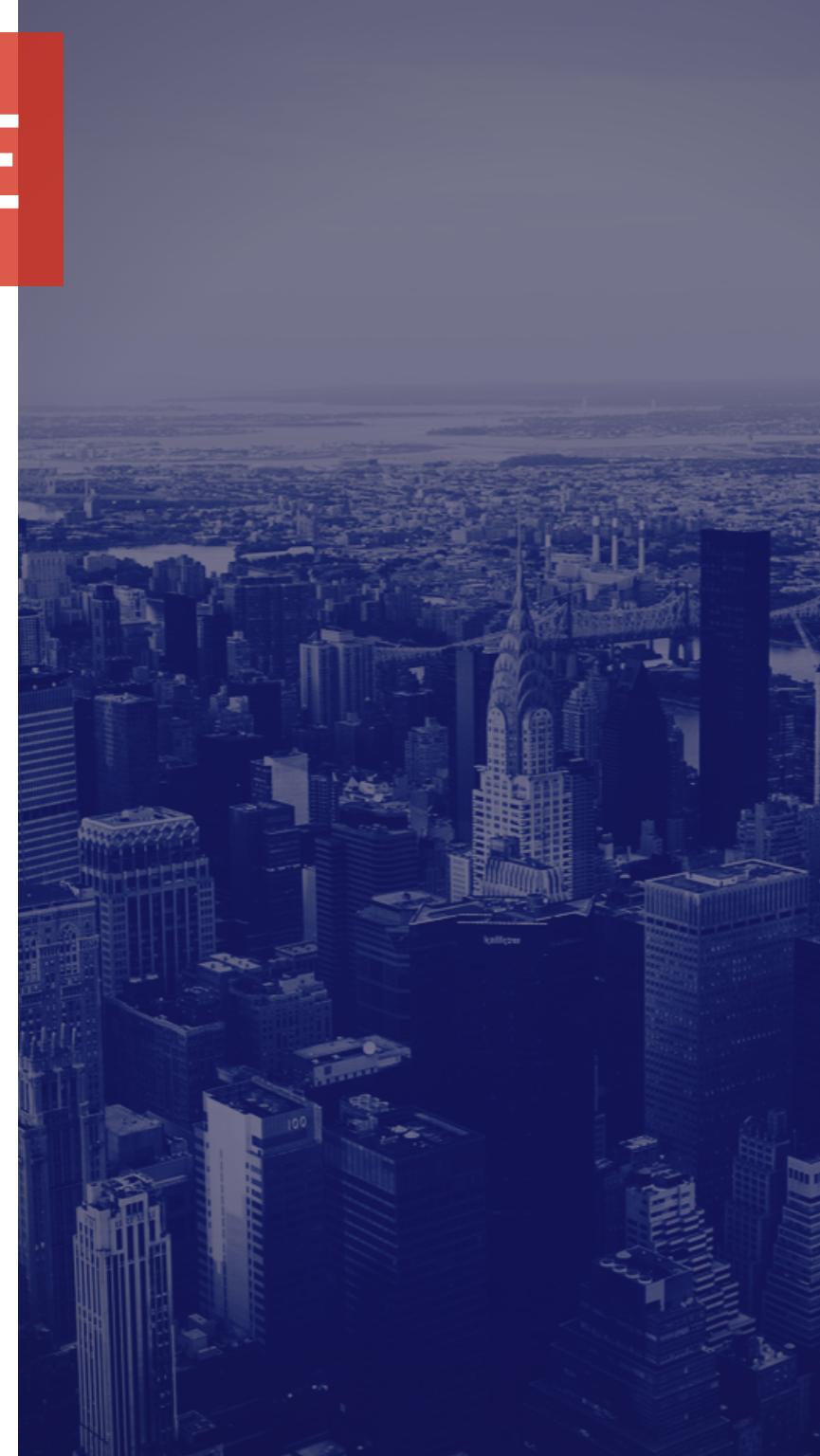
The Company is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures.

Borrowers can easily access **lower interest rate loans** through a fast online interface.

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***

Borrowers who default cause the largest amount of loss to the lenders are labelled as '**charged-off**' are the '**defaulters**'.



# GRAMENER

## CASE STUDY

Anurag Maheshwari  
Anupam Rajput  
Subrata Sahu

# THE BUSINESS OBJECTIVE



To identify risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss

Identification of such applicants using EDA

The company can utilize this knowledge for its portfolio and risk assessment

**DATA COLLECTION**

**DATA CLEANING**

**CONCLUSION**

**DATA ANALYSIS**





# DATA CLEANING

## IDENTIFYING

- NA values
- emp\_length have 1075 'n/a' values
- emp\_length contain 2.7% of 'n/a' values
- revol\_util contain 0.13% of total data
- Removing corresponding rows
- loan\_status = 'Charged Off'

We identified 16 major columns which factors of interest

- loan\_amount
- funded\_amount
- annual\_inc term
- funded\_amount\_inv
- inq\_last\_6mths
- open\_acc
- verification\_status
- emp\_length
- installment
- int\_rate
- home\_ownership
- dti
- purpose
- grade
- revol\_util



# ANALYSIS

## UNI-VARIATE ANALYSIS

Univariate analysis is the simplest form of analyzing data. “Uni” means “one” has **only one variable**. It takes data, summarizes that data and finds patterns in the data

## MULTI-VARIATE ANALYSIS

Bivariate analysis is one of the simplest forms of quantitative (statistical) analysis. It involves the analysis of **two or more variables** for the purpose of determining the correlation between them

## DRIVER VARIABLES

To understand the **driving factors** i.e. the driver variables, the variables which are strong indicators of default

# Uni-Variate Analysis

We recorded the skewness of each variable to check the actual distribution. The graphs are shown in the next few slides

## VARIABLE

## SKEWNESS

## INFERENCE

loan\_amt

1.05

Slightly skewed towards the right

installment

1.12

Slightly skewed towards the right

open\_acc

1.00

Slightly skewed towards the right

revol\_util

-0.04

Close to normal distribution

annual\_inc

31.11

Strongly skewed towards the right

int\_rate

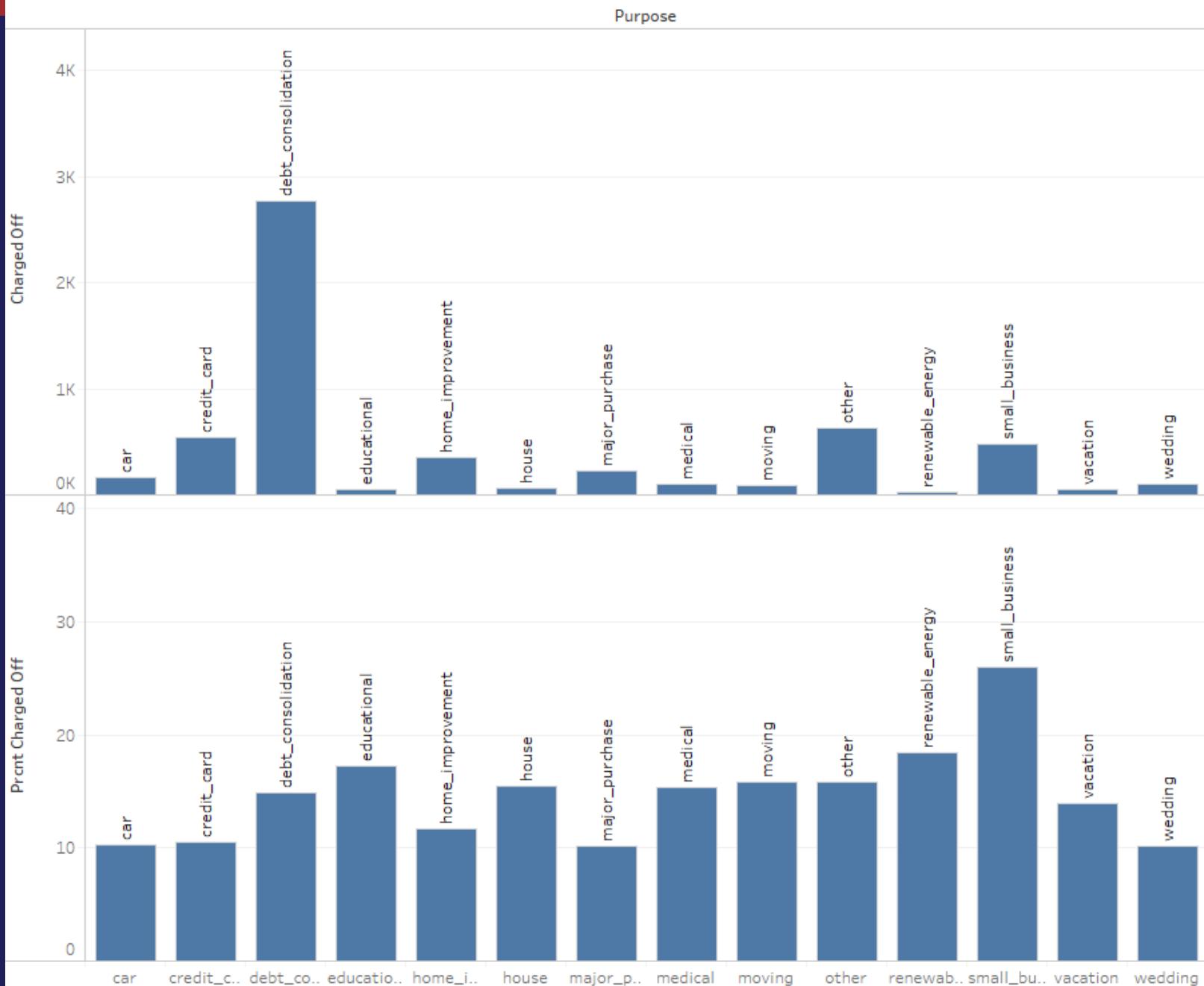
0.29

Close to normal distribution

# UNI- VARIATE ANALYSIS

## #01 PURPOSE ANALYSIS

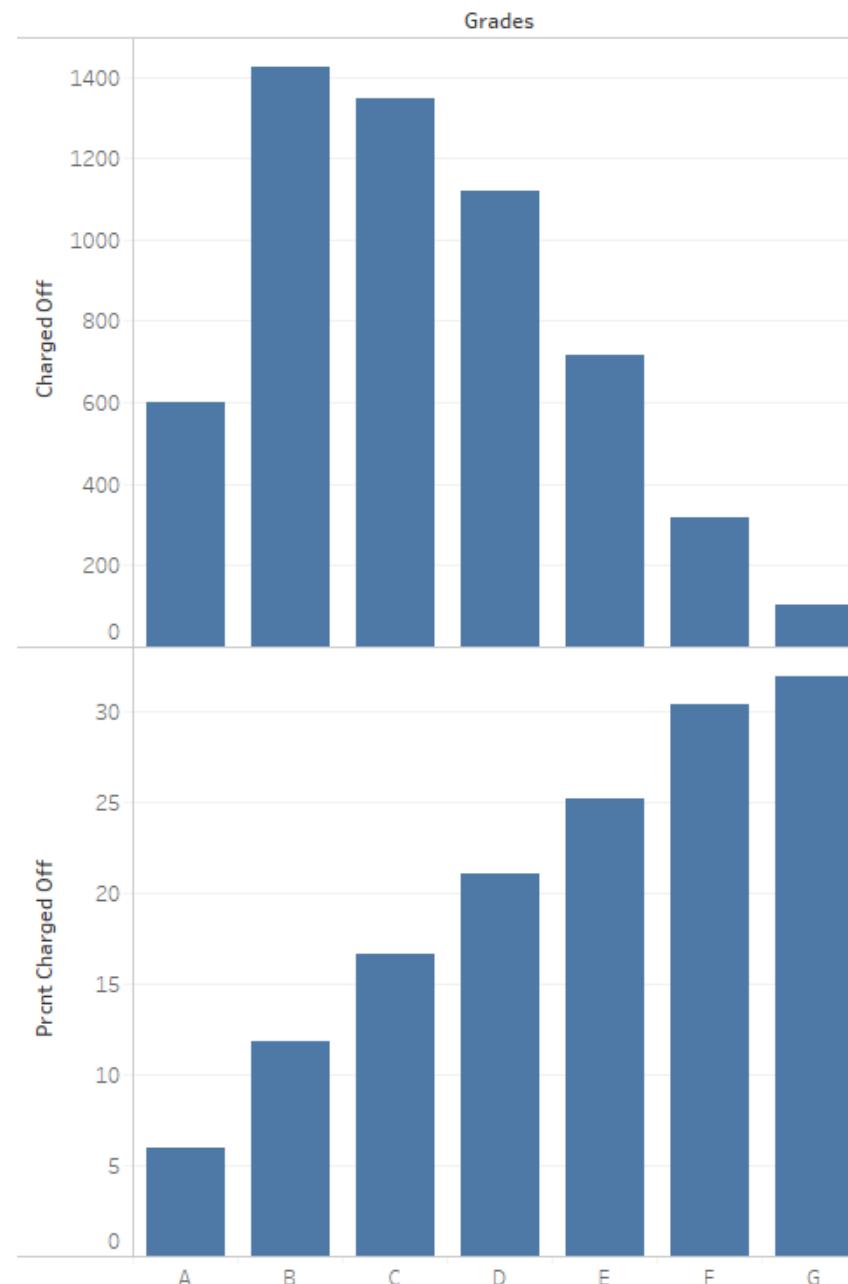
- The **Debt Consolidation** have the highest number of defaulters
- In case of percentage of loan defaulters among all the type of loan taken, **small business loans** tend to have the highest percent



# UNI- VARIATE ANALYSIS

## #02 GRADE ANALYSIS

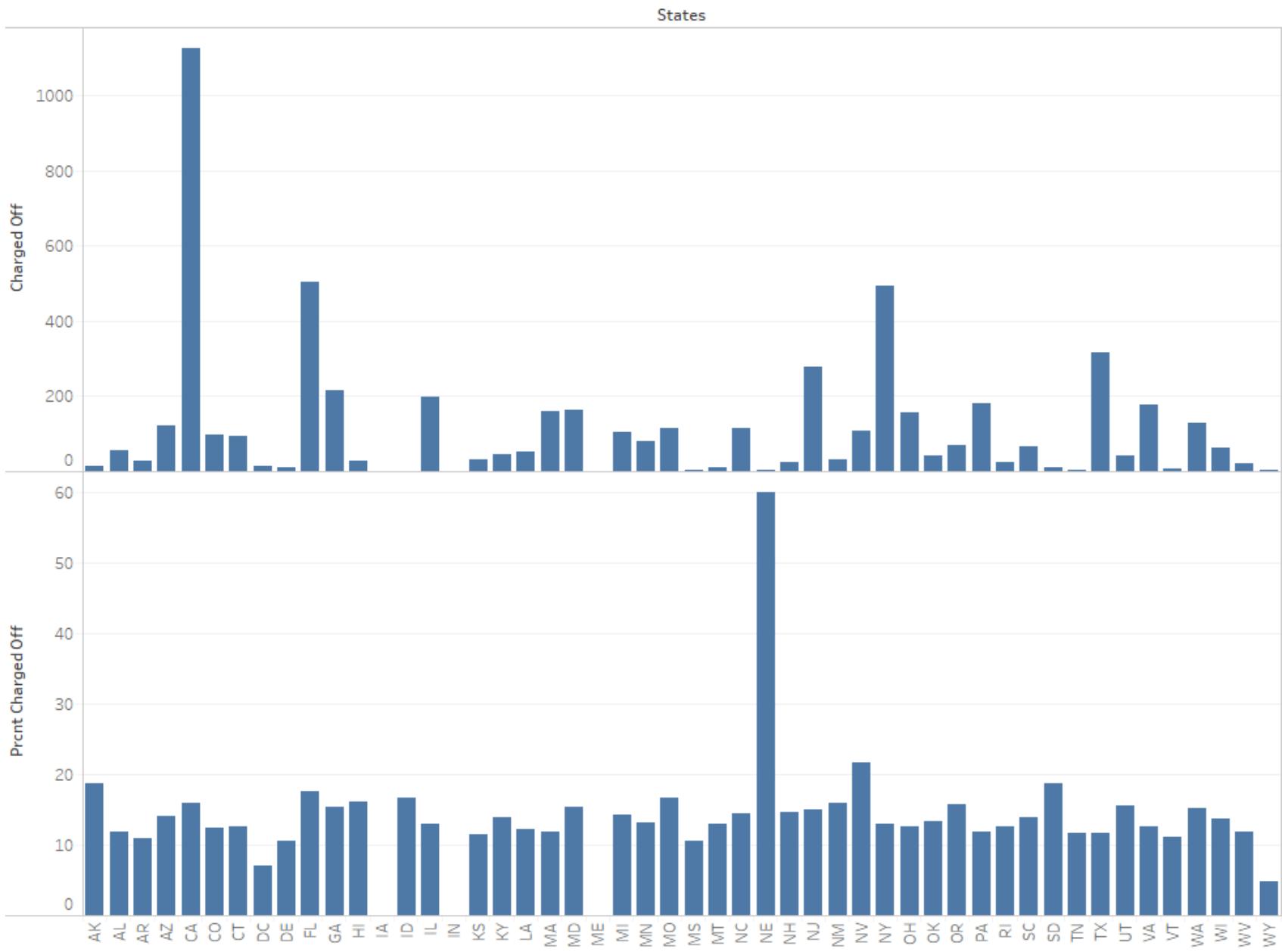
- Grade B loans tends to be the highest number of defaulters
- In case of percentage, Grade G (which is very small in number) tends to get high percentage of chance to get loan Charged Off



# UNI- VARIATE ANALYSIS

## #03 STATE ANALYSIS

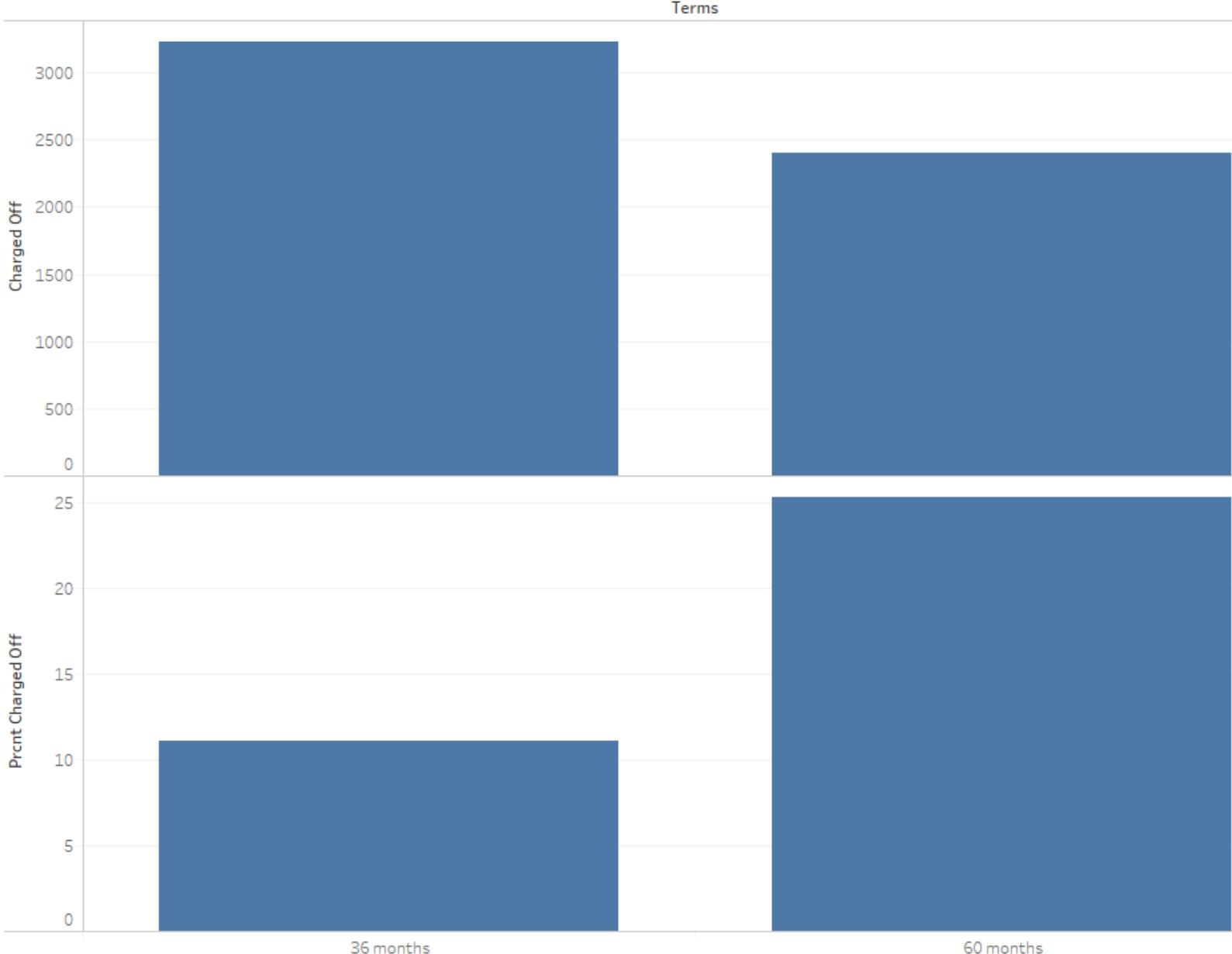
- CA state has the highest number of loans
- A spike at NE shows that even with small number of loans, NE state have almost all their loans as Charged Off



# UNI- VARIATE ANALYSIS

## #04 TERM ANALYSIS

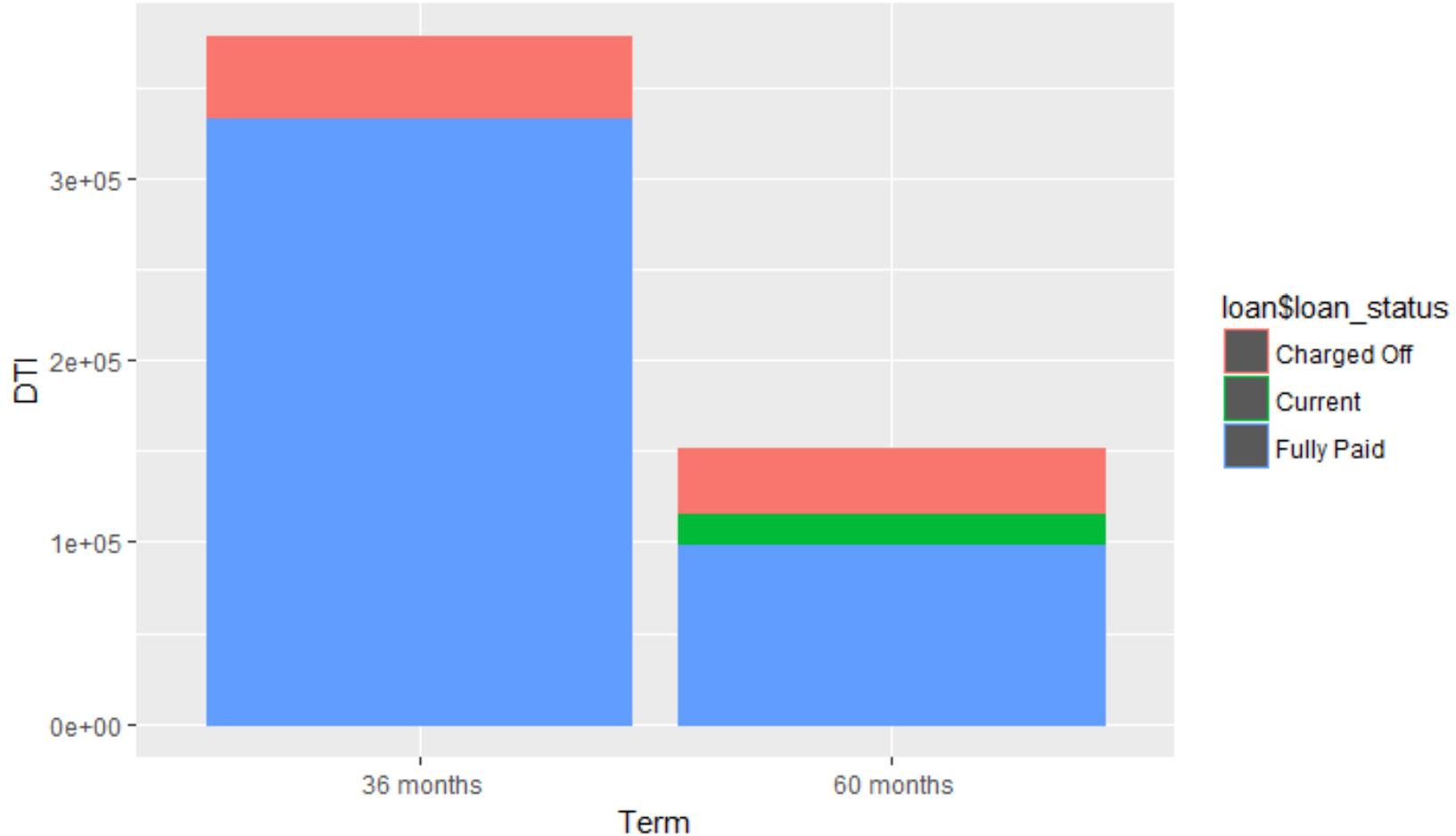
- It is an interesting insight
- The graph below shows that number of loans being charged off are **slightly higher for 36 month term loans.**
- Also the tendency of loans to get defaulted is **very high in case of 60 month term loans**



# UNI- VARIATE ANALYSIS

## #04 TERM ANALYSIS

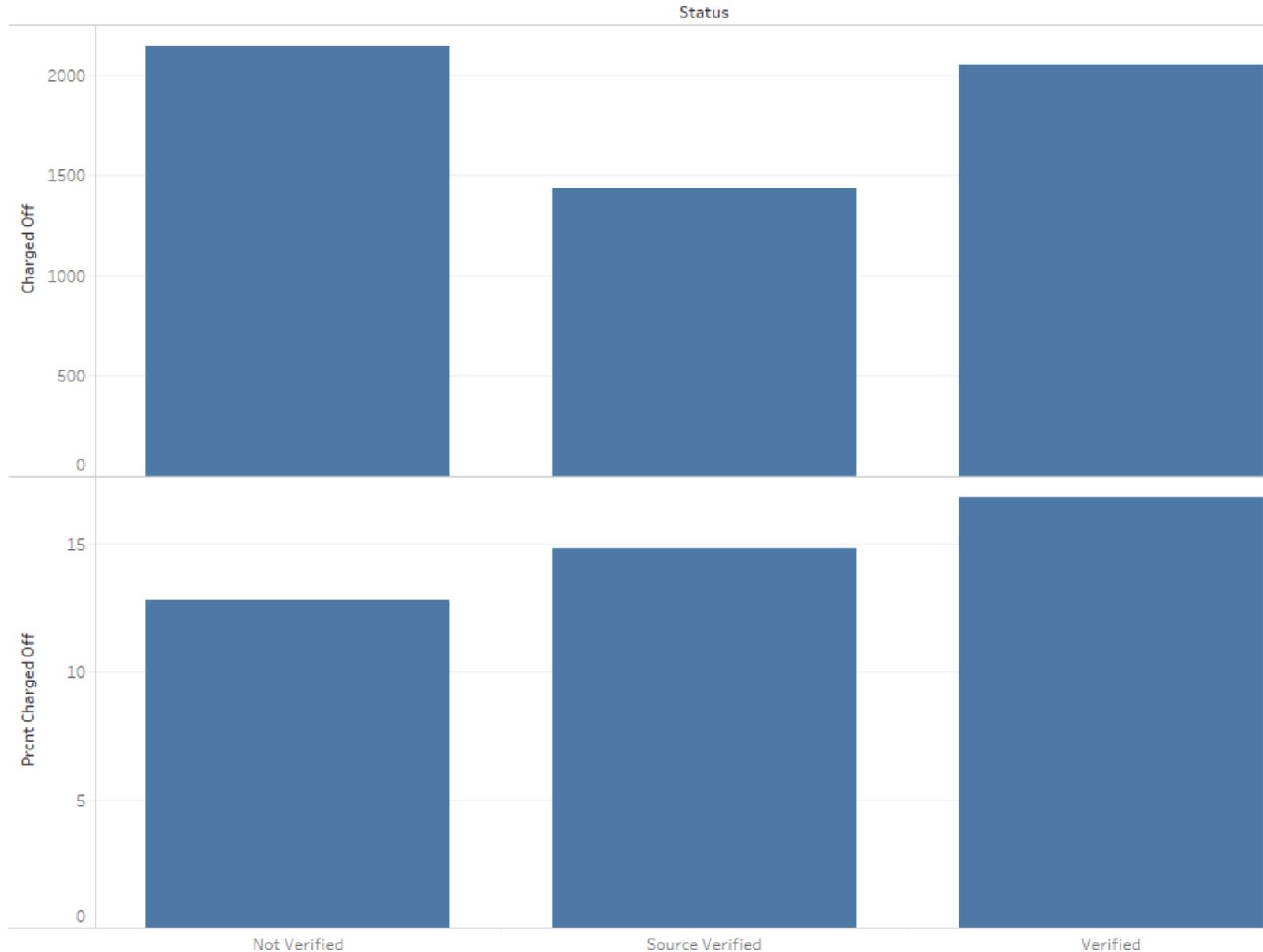
- There are **NO Current Loans** for 36 month term



# UNI- VARIATE ANALYSIS

## #05 VERIFICATION STATUS ANALYSIS

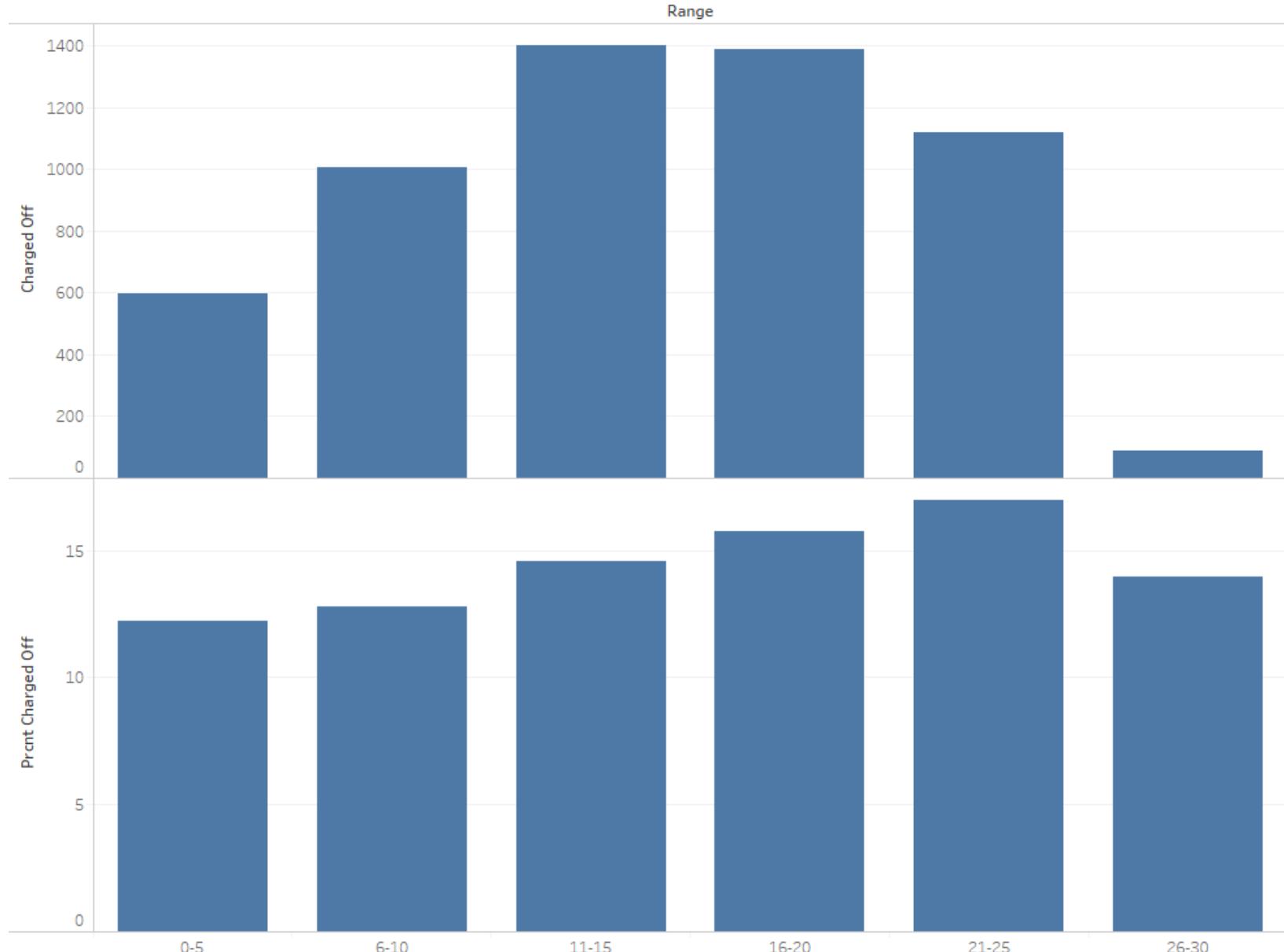
- The graph depicts that those who are **not verified** are **very large** in number of getting their loans defaulted in comparison to those who get verified by LC and their source verified.
- The percentage graph shows that **tendency to get defaulted is highest** for those who are **verified** in comparison to those who are not verified and source verified.



# UNI- VARIATE ANALYSIS

## #06 DTI ANALYSIS

- It is an important field of analysis which measures an individual's ability to manage monthly payments and repay debts
- A derived metric (range metric) to analyse the DTI trend shows that those having
  - DTI **11-20%** are high in number of loan defaulters
  - DTI in range of **21-25%** are most probable to get their loans charged off

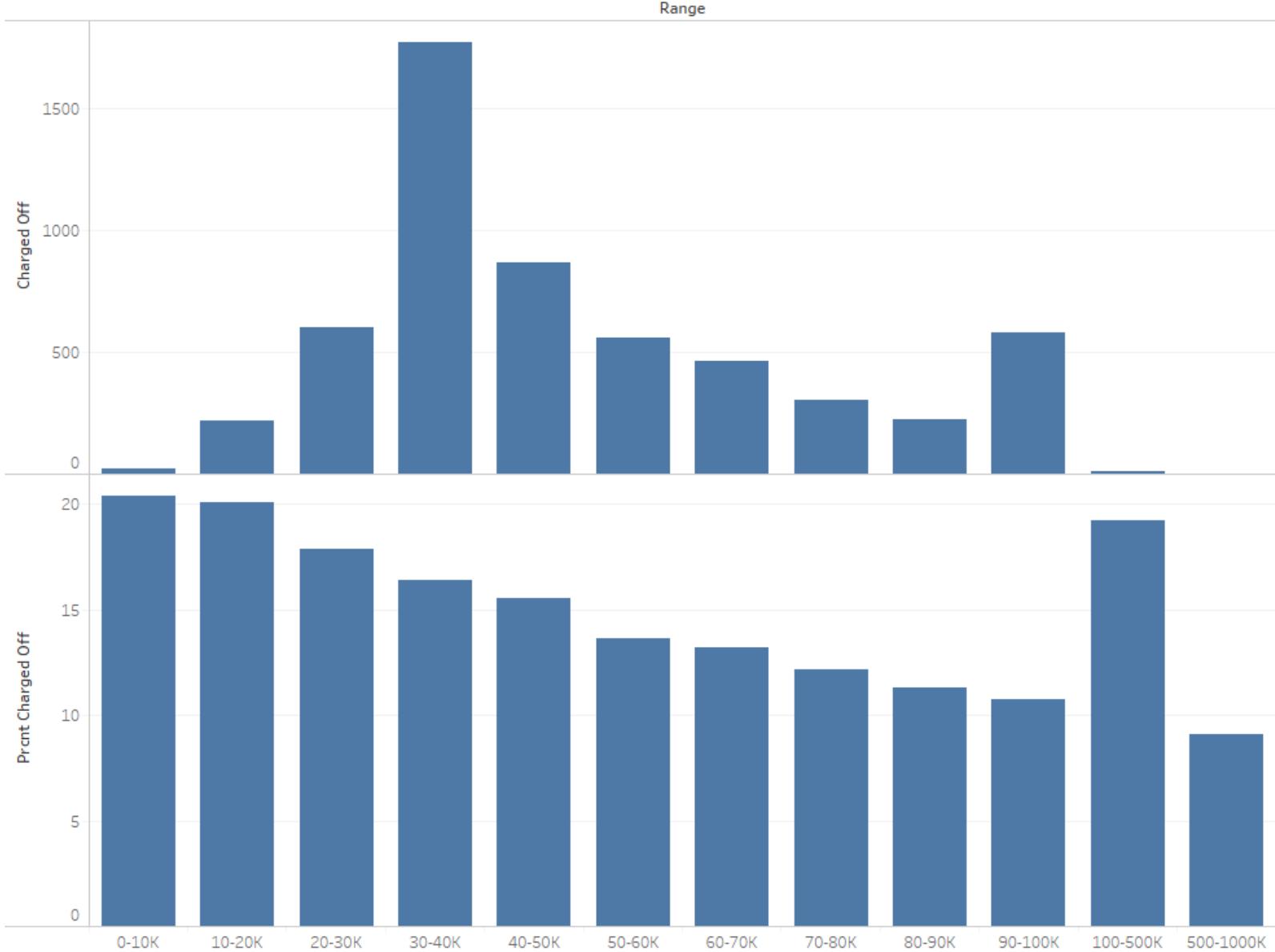


# UNI- VARIATE ANALYSIS

## #07 ANNUAL INCOME ANALYSIS

On basis of a derived metric which is a range metric on annual income, we can infer 3 things out of it

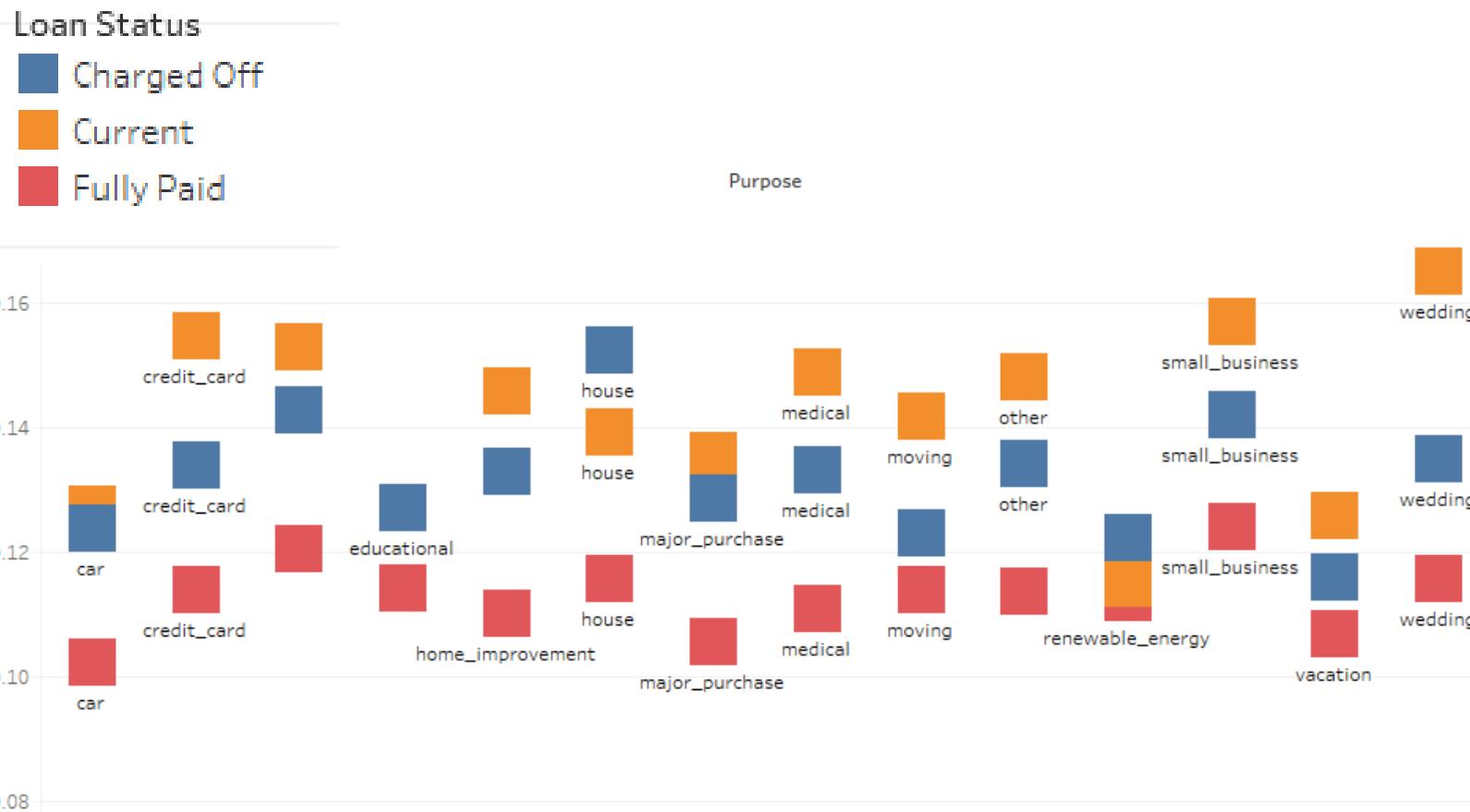
- **Very low income (10-20K)** tend to have high probability of getting **loan defaulted** for obvious reason
- **Income in range of 30-40k** are the highest in number of getting their **loan charged off**
- **Very rich (income 100-500K)** tend to have high probability of being **loan defaulters**



# MULTI-VARIATE ANALYSIS

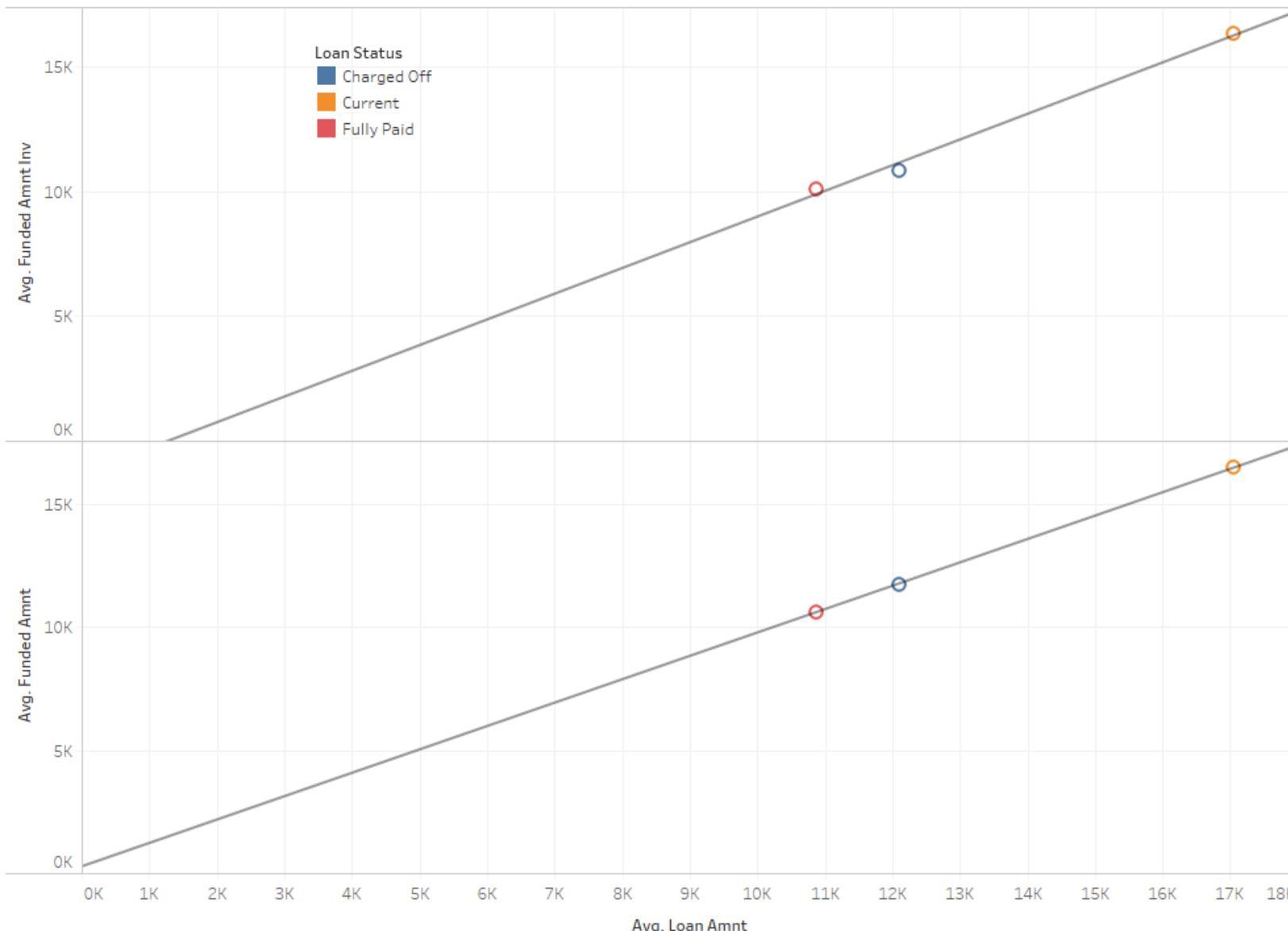
- The average interest rate for the loan defaulters is higher than the average interest rate for Fully paid
- The average **interest rate for the current is highest**. The reason is recovery because of which the interest rate have been continuously increasing
- In case of housing loans, the gap between **avg interest rate of defaulters and fully paid is very high**.
- Contrasting to others, the **avg rate for current is lower than the defaulters purging temporary defaulters**

## #01 INTEREST RATE VS PURPOSE



# MULTI-VARIATE ANALYSIS

- All three are **highly correlated** measure
- Except **funded\_amnt\_inv** shows **slightly less correlation** for Charged Off than for Fully Paid and Current
- Correlation **funded\_amnt\_inv** (agreed by investors) with loan amount applied should be **greater than or equal to 95%**



#02

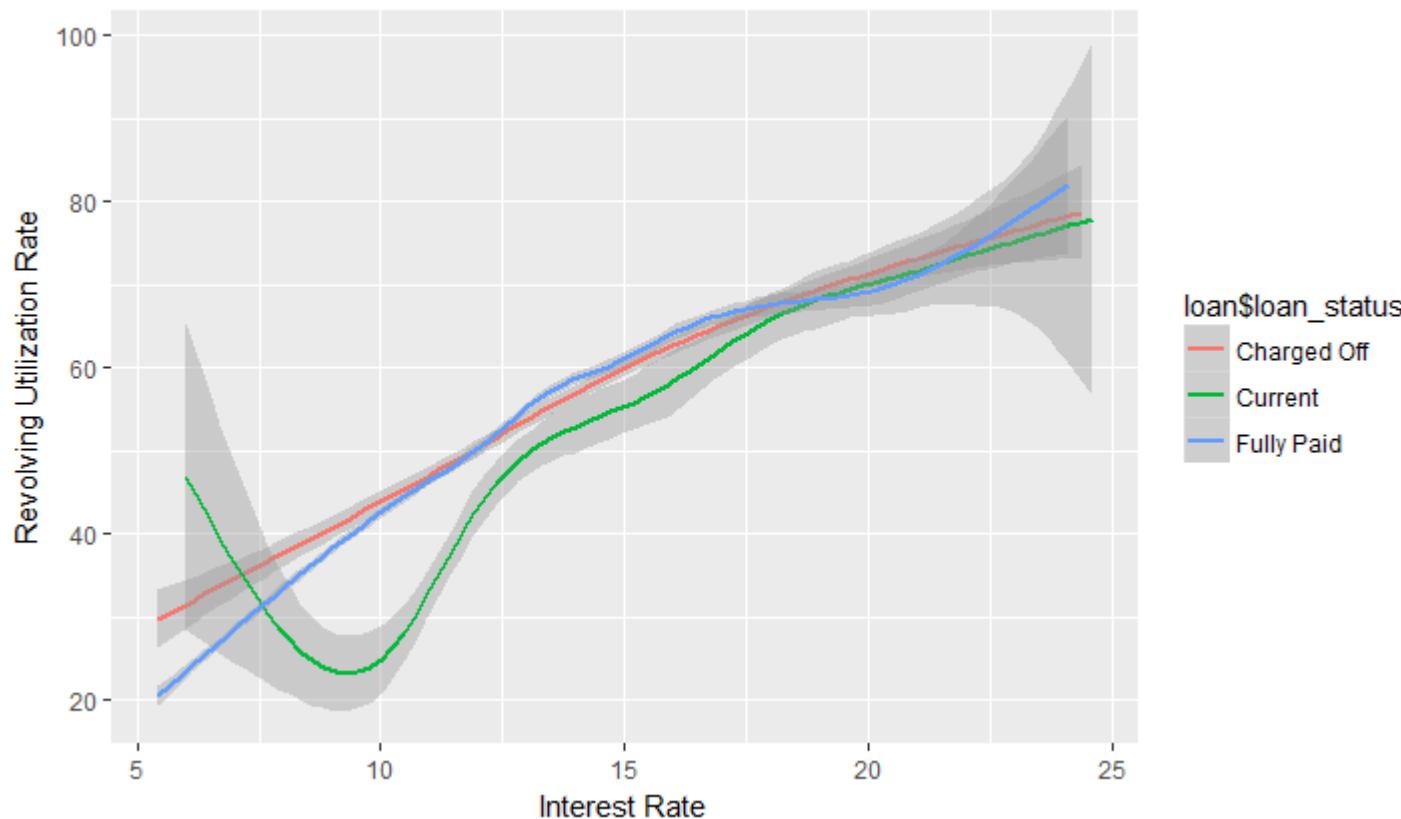
## Correlation between

- Loan Amount Applied by the borrower
- Total Amount committed by investors
- Total Amount committed to that Loan at that Time

# MULTI-VARIATE ANALYSIS

- Revolve Utilization rate is an important factor and **contribute almost 30% in calculating credit score**
- To Maximize credit score, we want Revolving Utilization Rate as low as possible
- Thus we can deduce:
  - There is **quite difference in the correlation between charged off and Fully Paid**
  - At interest rate **0-10% Utilization Rate for current loans decreases** which is contrasting to other loan status

*We conclude that 5-10% interest rate is an interest rate to which LC can consider for applications having Low Revolving Utilization Rate*



#03

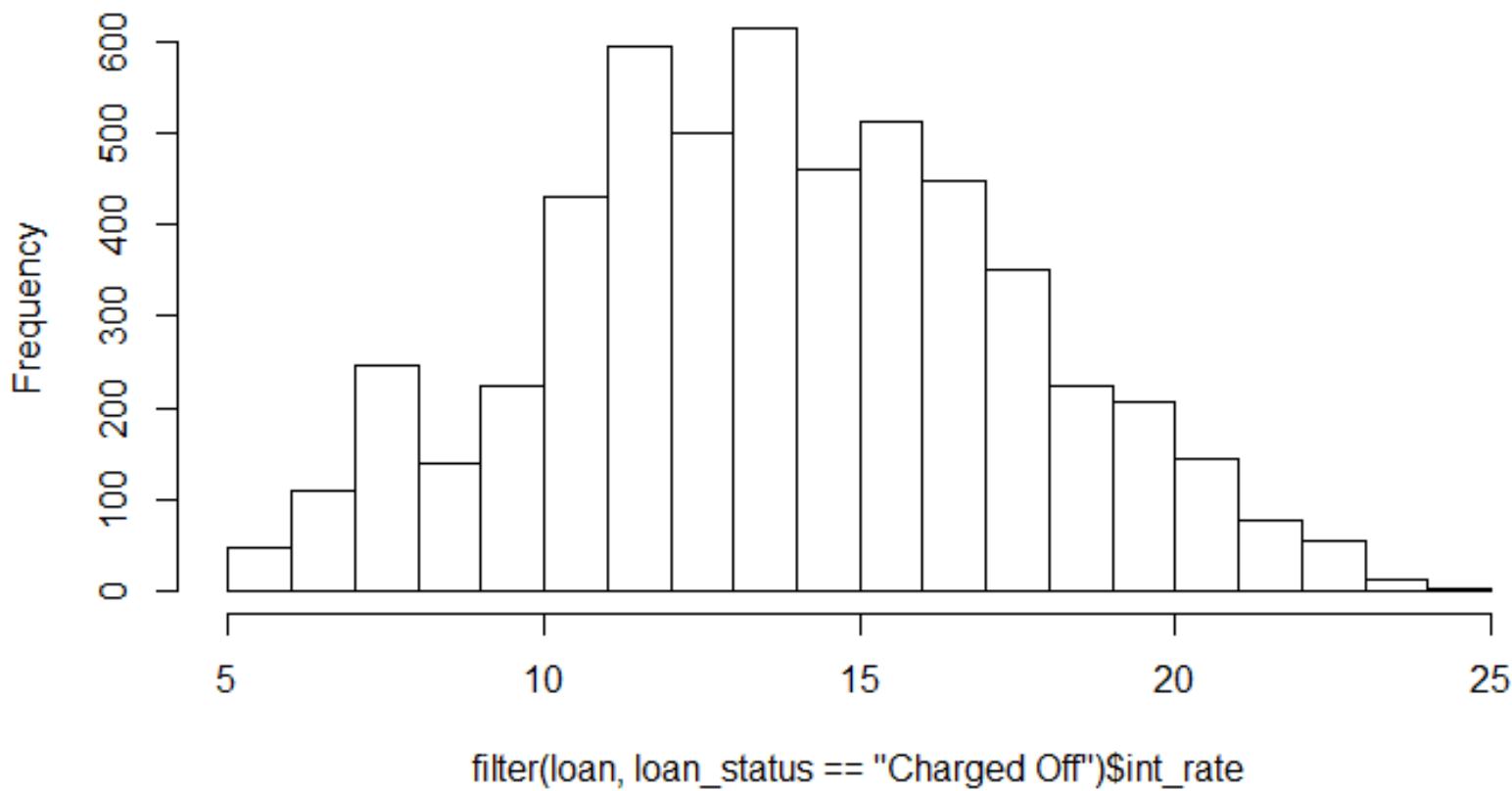
## Correlation between

Interest  
Rate

- Revolving Utilization Rate

- For Charged Off loan interest rate is higher for 11-13

Histogram of filter(loan, loan\_status == "Charged Off")\$int\_rate



#04

Analysis Of  
Rate of  
Interest  
with  
Various  
variables

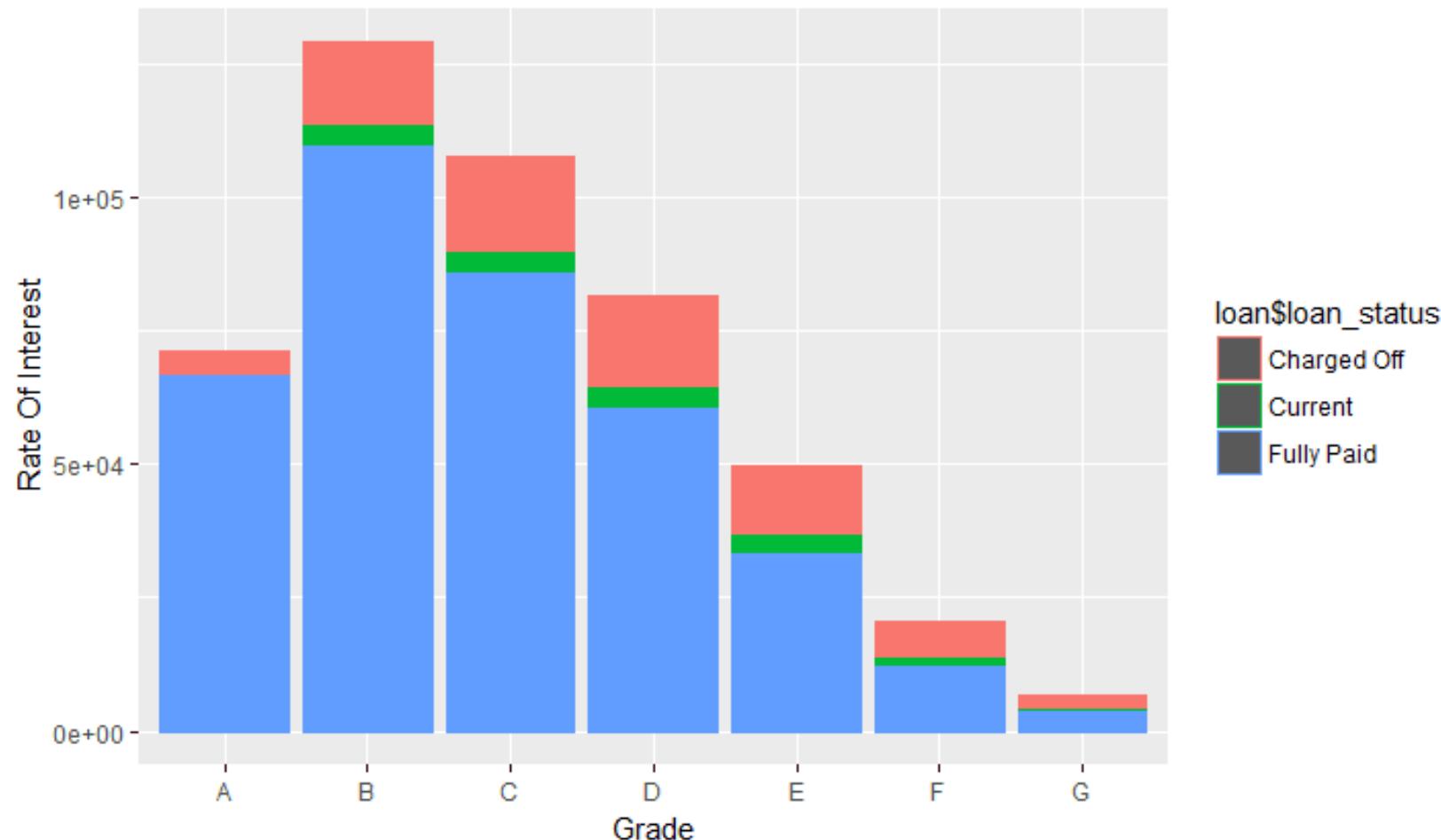
# MULTI-VARIATE ANALYSIS

#05

Rate Of Interest  
Vs  
Grade

- Grade B tends to be the higher Fully Paid borrowers for high interest of rate

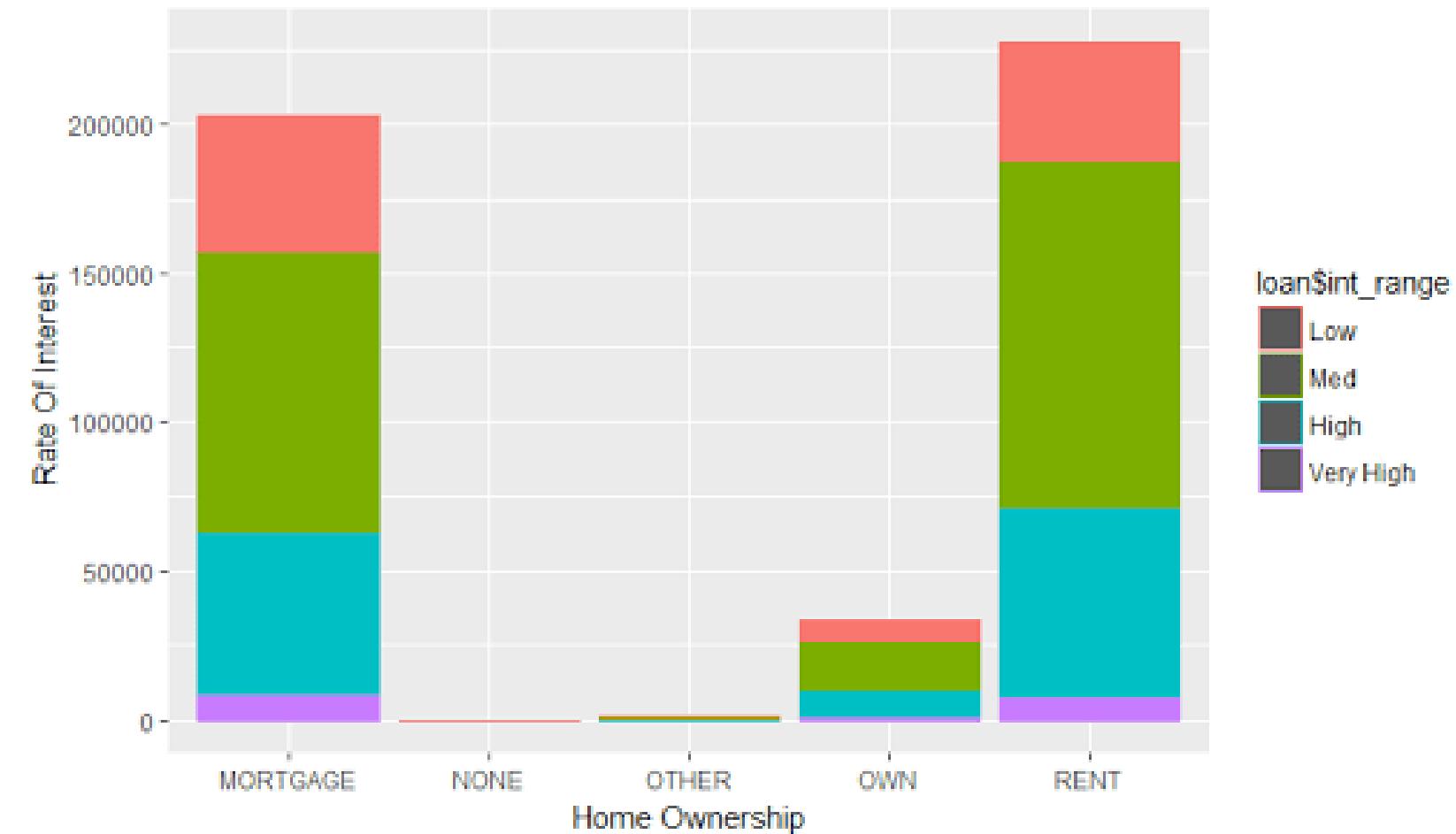
Rate Of Interest Vs Grade



# MULTI-VARIATE ANALYSIS

- Rent and Mortgage tends to have High tendency to Fully Paid the loan for High Interest

Rate Of Interest Vs Home Ownership



#06  
Interest Rate with Home Ownership

# CONCLUSION

On the basis of our uni and multi-variate analysis we conclude that following factors are the variables which are strong indicators of default

- loan\_amount
- funded\_amount
- annual\_inc term
- funded\_amount\_inv
- inq\_last\_6mths
- open\_acc
- verification\_status
- emp\_length
- installment
- int\_rate
- home\_ownership
- dti
- purpose
- grade
- revol\_util



THANKS  
FOR WATCHING