# GRAMENER CASE STUDY

Group Name:

1. Rajashekhar Battula
2. Satish Kumar
3. Darshna
4. Sarvan Kumar Yadav

# PROBLEM STATEMENT

## INTRODUCTION

- In this case study, real business problems are solved using EDA.
- A basic understanding of risk analytics in banking and financial services.
- Data is used to minimise the risk of losing money while lending to customers.

## BUSINESS UNDERSTANDING

### ❖ RISKS

- 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**, then approving the loan may lead to a **financial loss** for the company.

### ❖ COMPANY DECISION

- Loan Accepted:- (Fully Paid and Current) Non-Defaulter and (Charged-off ) Defaulter.
- Loan Rejected

## BUSINESS OBJECTIVE

- Aim of the case study to identify driving variables behind loan default.

# DATA CLEANING AND MANIPULATION

## ❖ POSSIBLE DATA INCONSISTENCIES

- Duplicate values of request ID.
- NA values in the columns.

## ❖ OTHER ISSUES

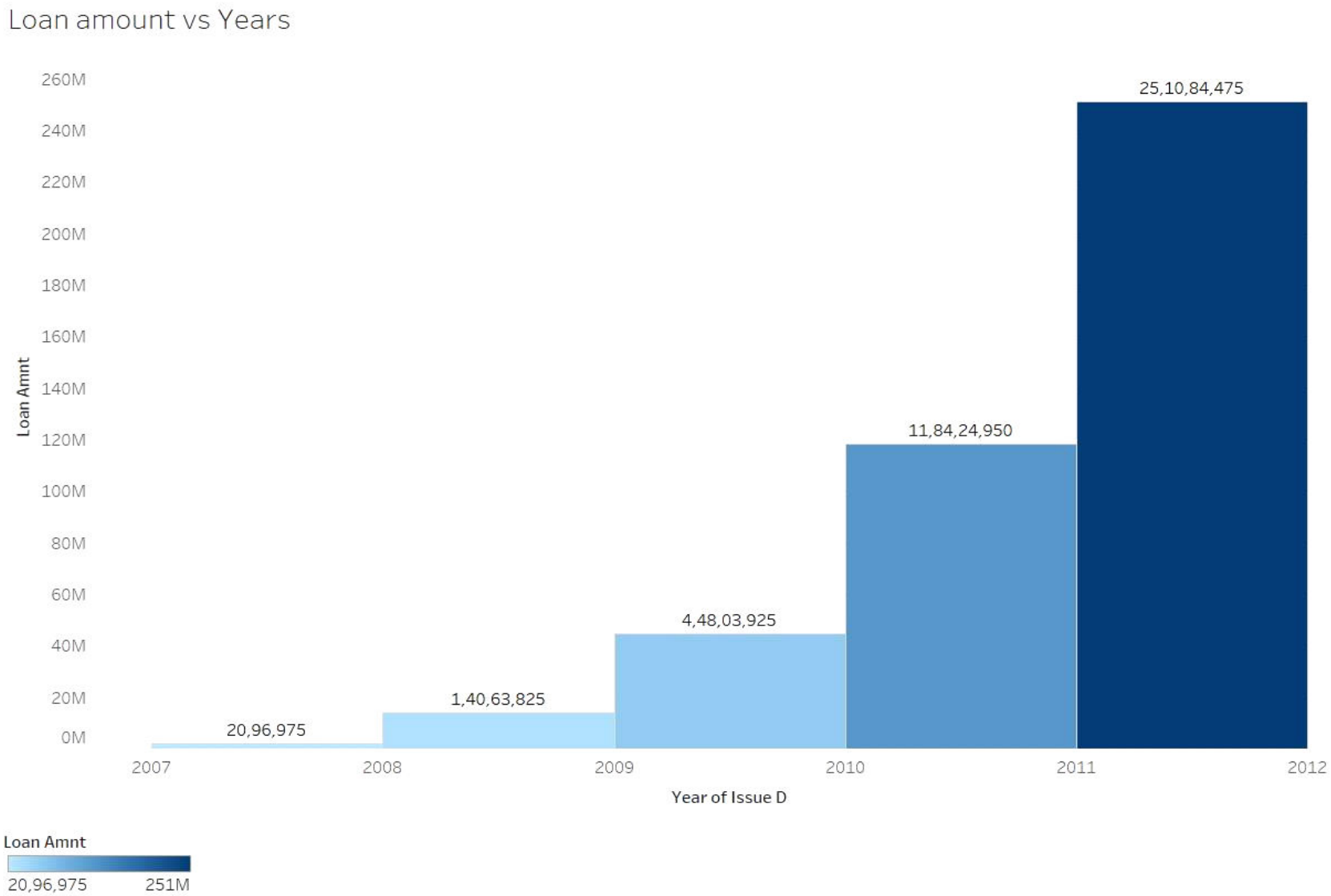
- Convert issue_d into datetime format.
- funded_amnt_inv and installment columns value round to 2 decimal places.
- Strip '%' from ‘int_rate’ column and values of the column round to 2 decimal places.
- Strip '%' from ‘revol_util’ column and values of the column round to 2 decimal places.
- Column emp_length replace '< 1' with 0 and strip the sub string '+' and ‘years’.
- Impute employee length with mean and values of the column round to 2 decimal places.
- Remove 'xx' from 'zip_code‘.
- Combine ‘zip_code’ and 'addr_state' to one column as 'address‘.

# METHODOLOGY USED FOR ANALYSIS

- Univariate and segmented univariate analysis is done correctly and appropriate realistic assumptions are made wherever required. The analyses successfully identify the 5 important driver variables (i.e. Variables which are strong indicators of default).
- Business-driven, type-driven and data-driven metrics are created for the important variables and utilised for analysis. The explanation for creating the derived metrics is mentioned and is reasonable
- Bivariate analysis is performed correctly and is able to identify the important combinations of driver variables. The combinations of variables are chosen such that they make business or analytical sense.

# YEARLY BREAKUP OF THE TOTAL LOAN AMOUNT

- In this plot, the yearly breakup of the loan is shown.
- As expected the loan amount is increasing year by year. It has minimum at 2007, about 2 millions and maximum at 2012 about 251 millions.

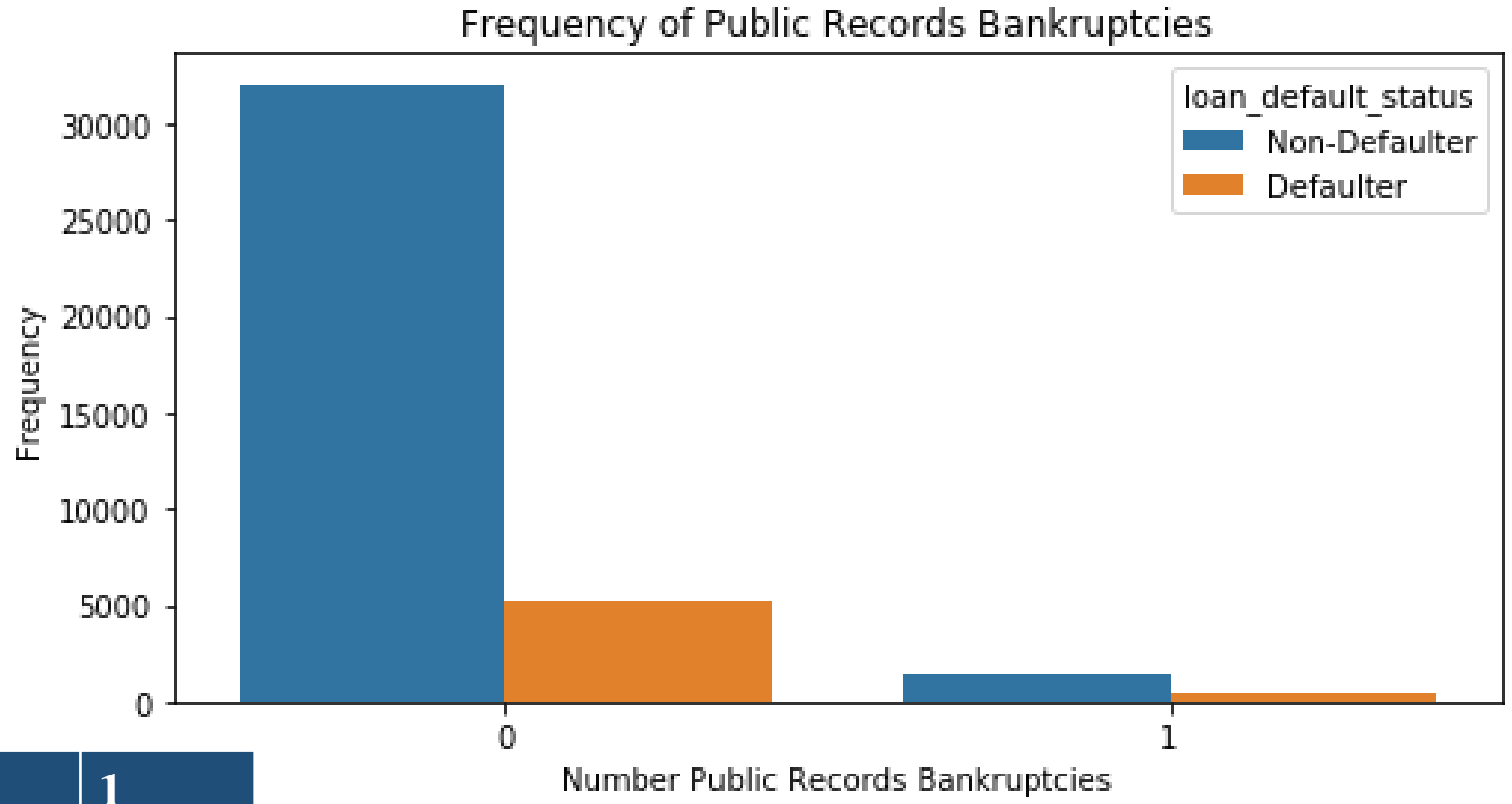


# ANALYSIS OF FACTORS BEHIND THE LOAN DEFAULTERS

## PUBLIC_RECORDS_BANKRUPTCIES VS DEFAULTER

- Percentage of charged off in not public bankrupt: 13.91%.
- Percentage of charged off in public bankrupt: 21.94%.

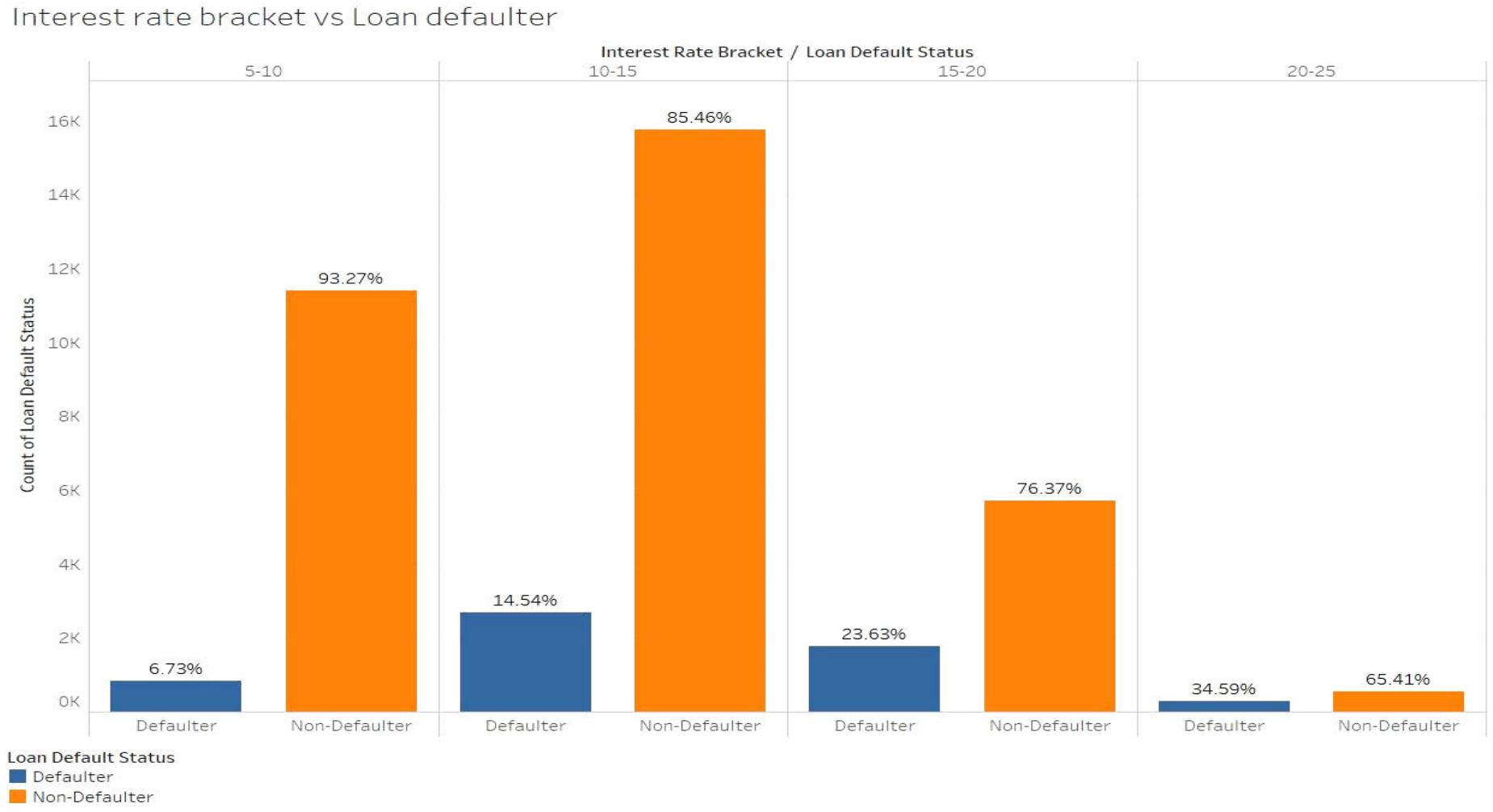
- Those person has pub_rec_bankruptcies record is more likely to default with 22% 0ver 14%




| Is_pub_rec_bankruptcies | 0 | 1 |
|---|---|---|
| Loan_default_status | | |
| Defaulter | 5183 | 366 |
| Non-Defaulter | 32069 | 1302 |

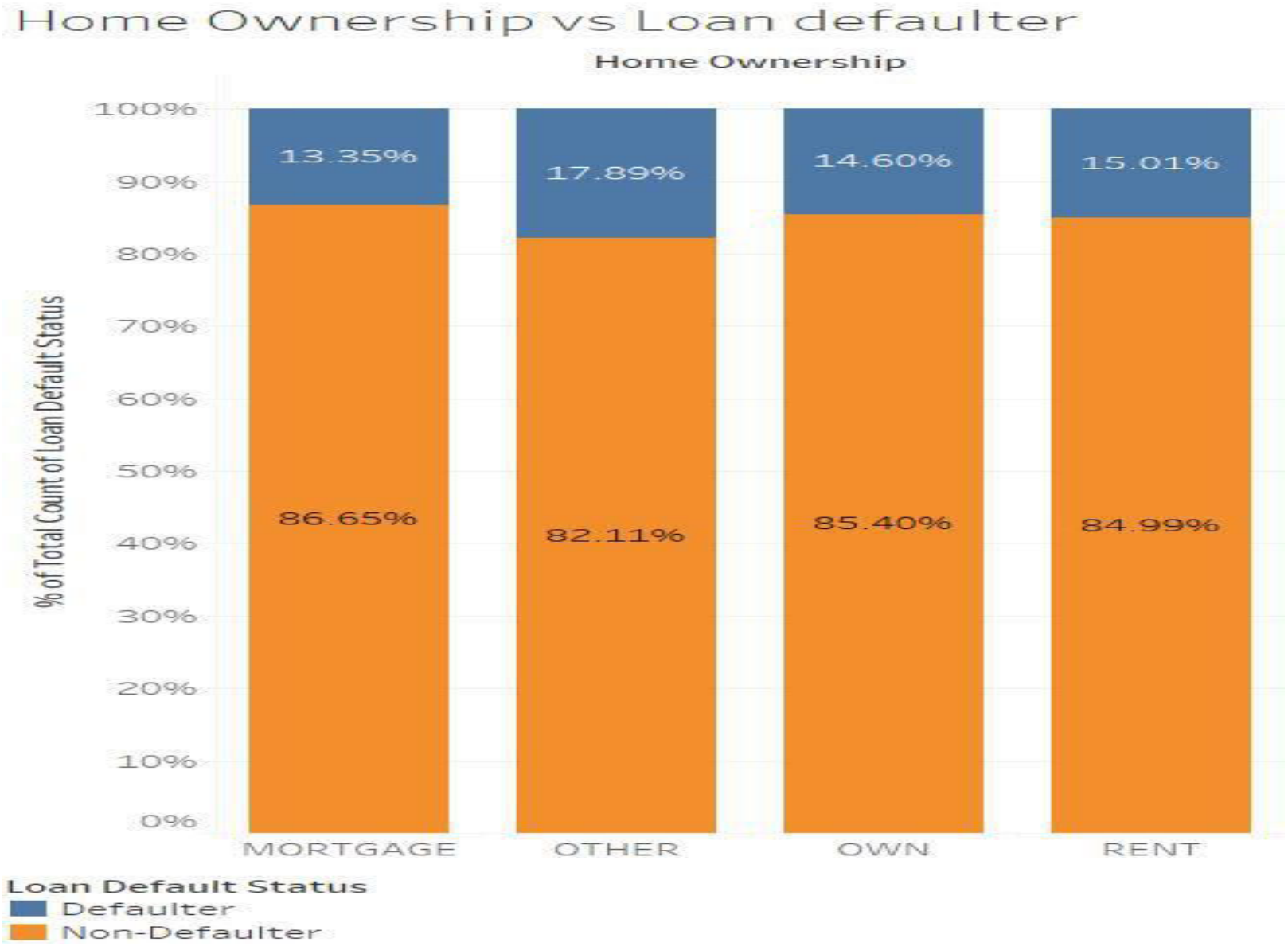
# INTEREST RATE VS DEFAULTER

In this plot, interest rate vs. defaulters plotted. Here quite clearly it shows, higher the interest more likely to default.



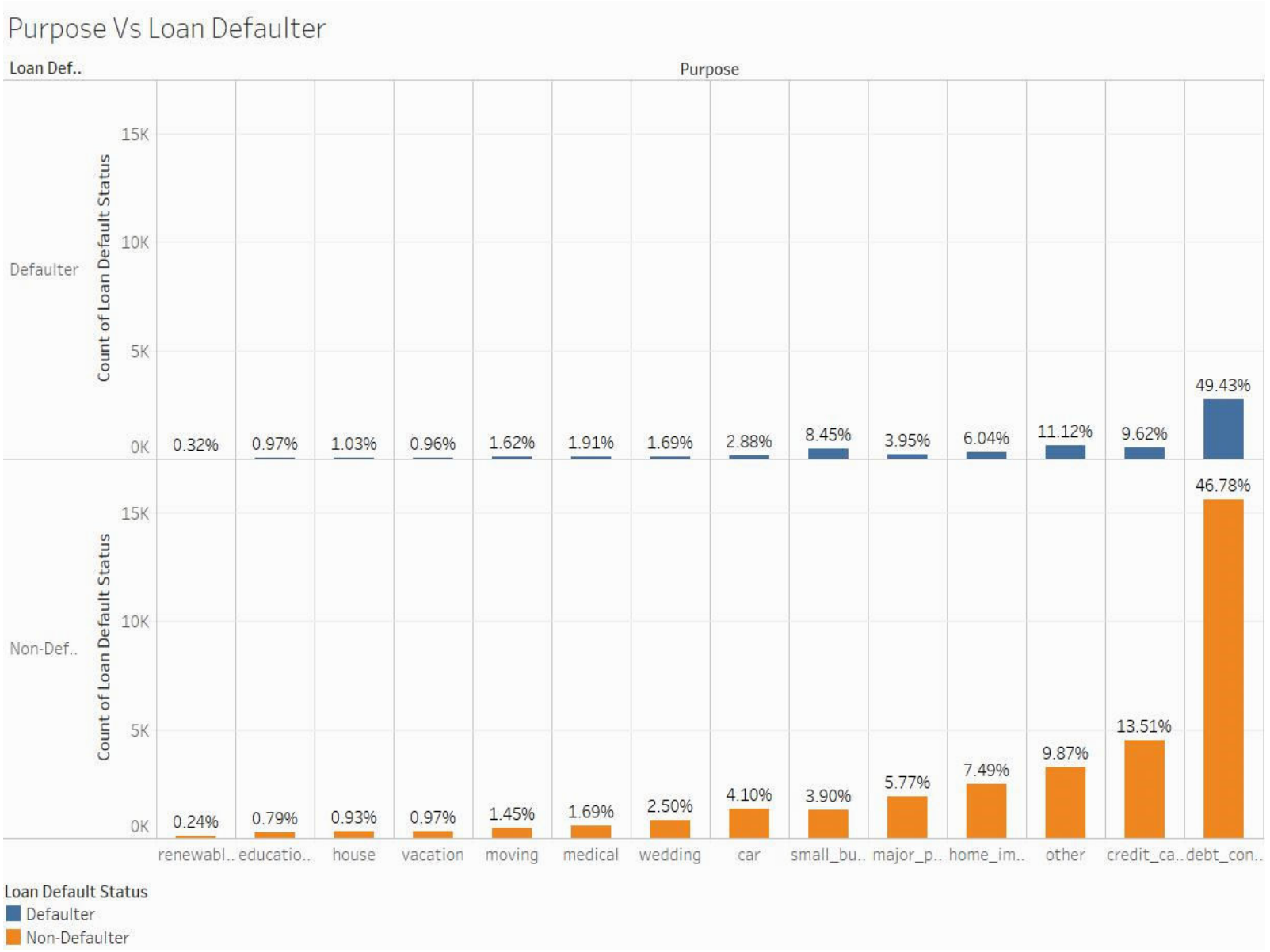
# HOME OWNERSHIP VS DEFAULTER

- This plot is showing relationship between home ownership and loan defaulters.
- Clearly Others and rented people are more likely defaulters.



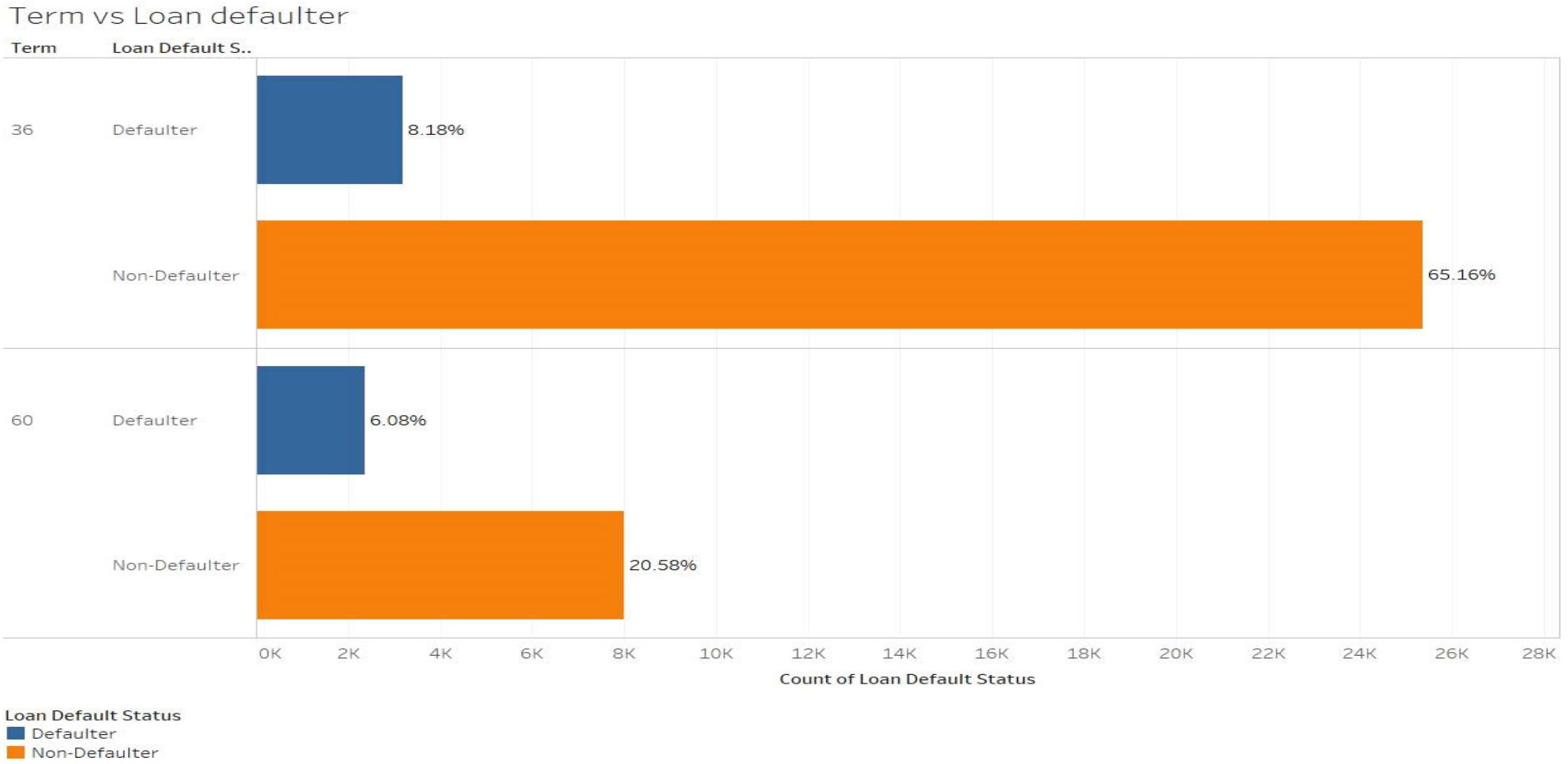
# PURPOSE VS DEFAULTER

- This plot is showing segmented plot.
- The top one showing percentage of defaulter vs. purpose of loan and bottom one showing percentage of non defaulter in each of the purpose.
- Clearly debt_consolidation, Credit card and others having higher no defaulters.



# TERM VS DEFAULTER

- This plot is showing term of the loans vs. loan status.
- Clearly, higher the terms higher is the no of defaulters.



# CONCLUSIONS

We have identified the five variables which is affecting the loan defaulters significantly. The variables are: Public Record of Bankruptcies, Interest rate, Home ownership, Purpose of loan and Loan Term.

- **pub_rec_bankruptcies**: Those person has pub_rec_bankruptcies record is more likely to default with 22% over 14% who has not.
- **Interest Rate**: Higher the interest rate higher the number of defaulters and so is the Grade and sub Grades. Grades and sub Grades are highly correlated.
- **Home ownership**: The customers whose home ownership were Rent and Others are most likely to do loan default.
- **Purpose**: The purpose of the loan is also found to be significant drivers for the customer to be defaulter. The debt_consolidation, credit card and others are having more no of defaulters
- **Term**: This also has positive correlation with the no of defaulter. That is higher the loan term (60 months) higher is the probability that the person default.

Although, we have found five most significant factors determining the loan defaulters in advance. We need to very careful in dealing with the customers as the correlation are not very strong between the loan status and any one of the variables therefore look at all the factors together not a single one to decide for the loan candidate.