Loan Lending Case Study

Business Understanding

You work for a consumer finance company which specialises in lending various types of loans to urban customers. When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:

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

The data given below contains the information about past loan applicants and whether they '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.

In this case study, you will use EDA to understand how consumer attributes and loan attributes influence the tendency of default.
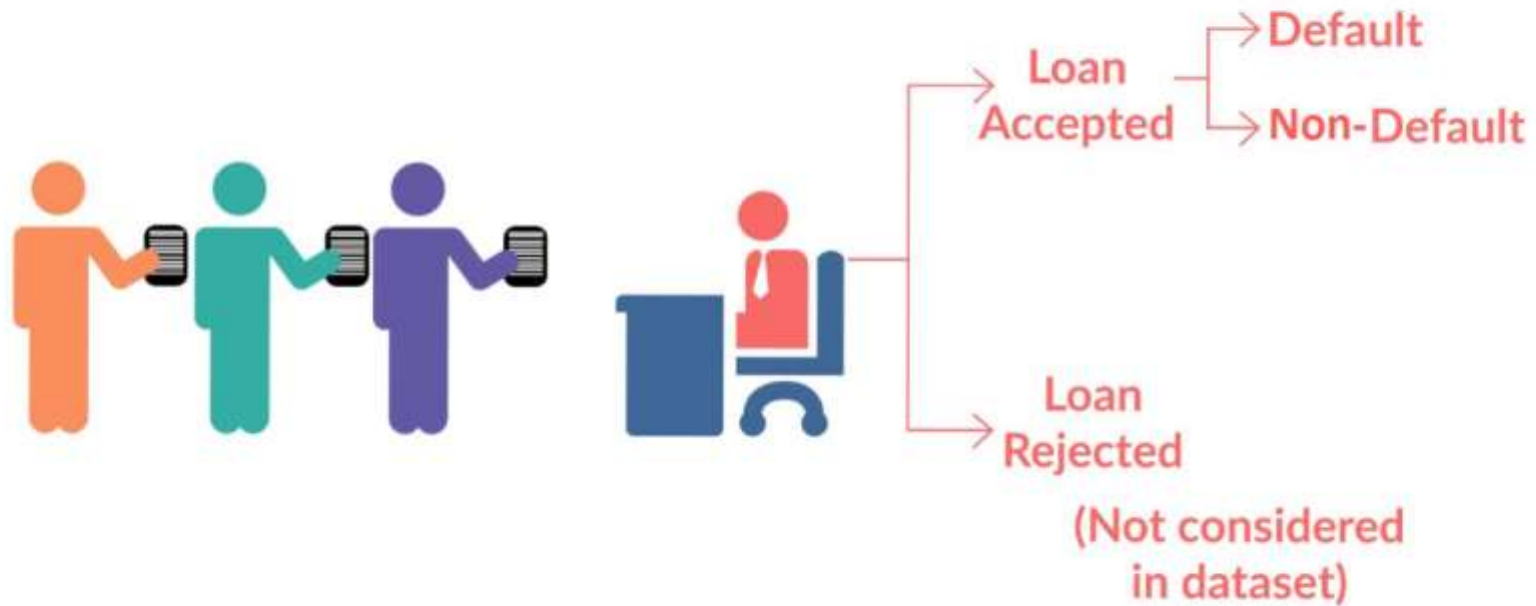
LOAN DATASET

Figure 1. Loan Data Set

Figure 1. Loan Data Set

When a person applies for a loan, there are two types of decisions that could be taken by the company:

Loan accepted : If the company approves the loan, there are 3 possible scenarios described below:

- Fully paid: Applicant has fully paid the loan (the principal and the interest rate)
- Current: Applicant is in the process of paying the instalments, i.e. the tenure of the loan is not yet completed. These candidates are not labelled as 'defaulted'.
- Charged-off: Applicant has not paid the instalments in due time for a long period of time, i.e. he/she has defaulted on the loan

Loan rejected : The company had rejected the loan (because the candidate does not meet their requirements etc.). Since the loan was rejected, there is no transactional history of those applicants with the company and so this data is not available with the company (and thus in this dataset)

Business Objectives

This 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.

Like most other lending companies, 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. In other words, borrowers who default cause the largest amount of loss to the lenders. In this case, the customers labelled as 'charged-off' are the 'defaulters'.

If one is able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. Identification of such applicants using EDA is the aim of this case study.

In other words, 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 utilise this knowledge for its portfolio and risk assessment

Data understanding

Understanding data in term of business understanding We categorize our data in three types of variables below and consider below variables as our basis of analysis.
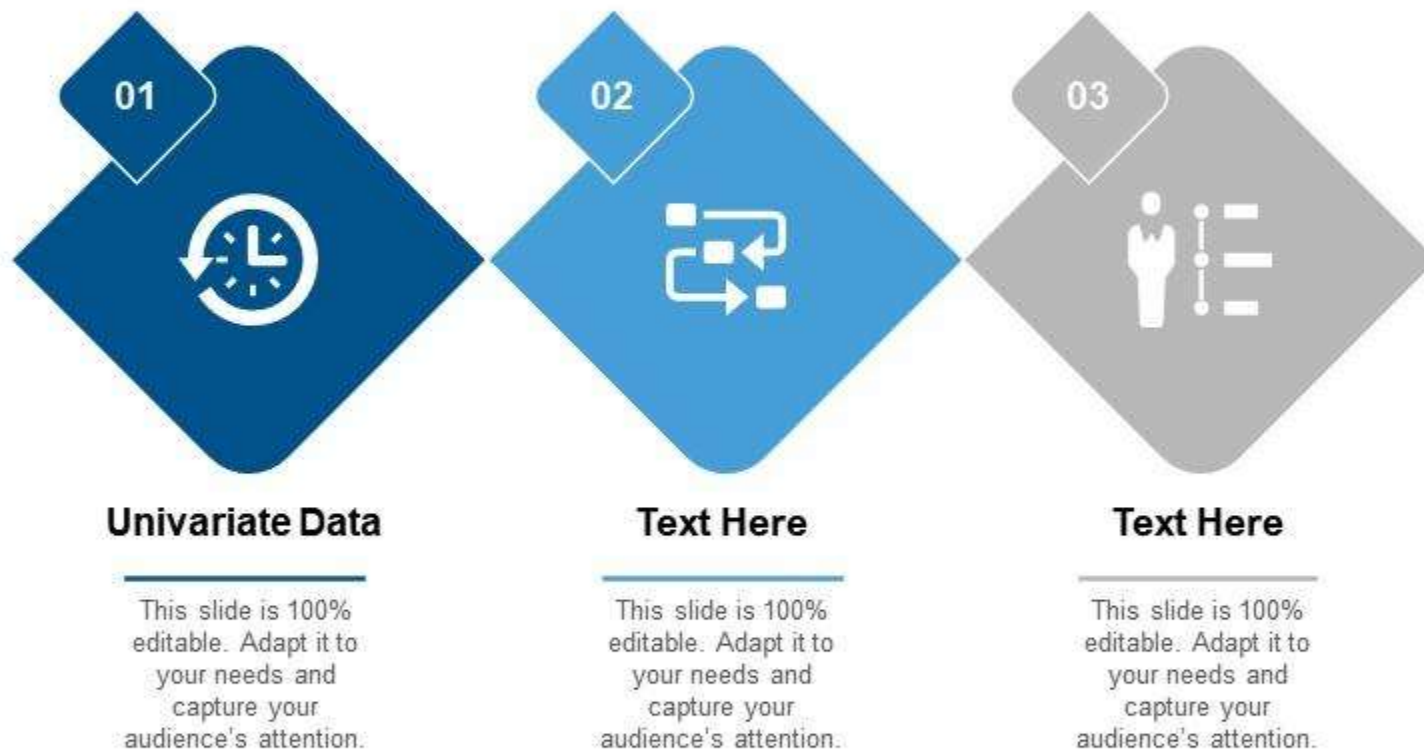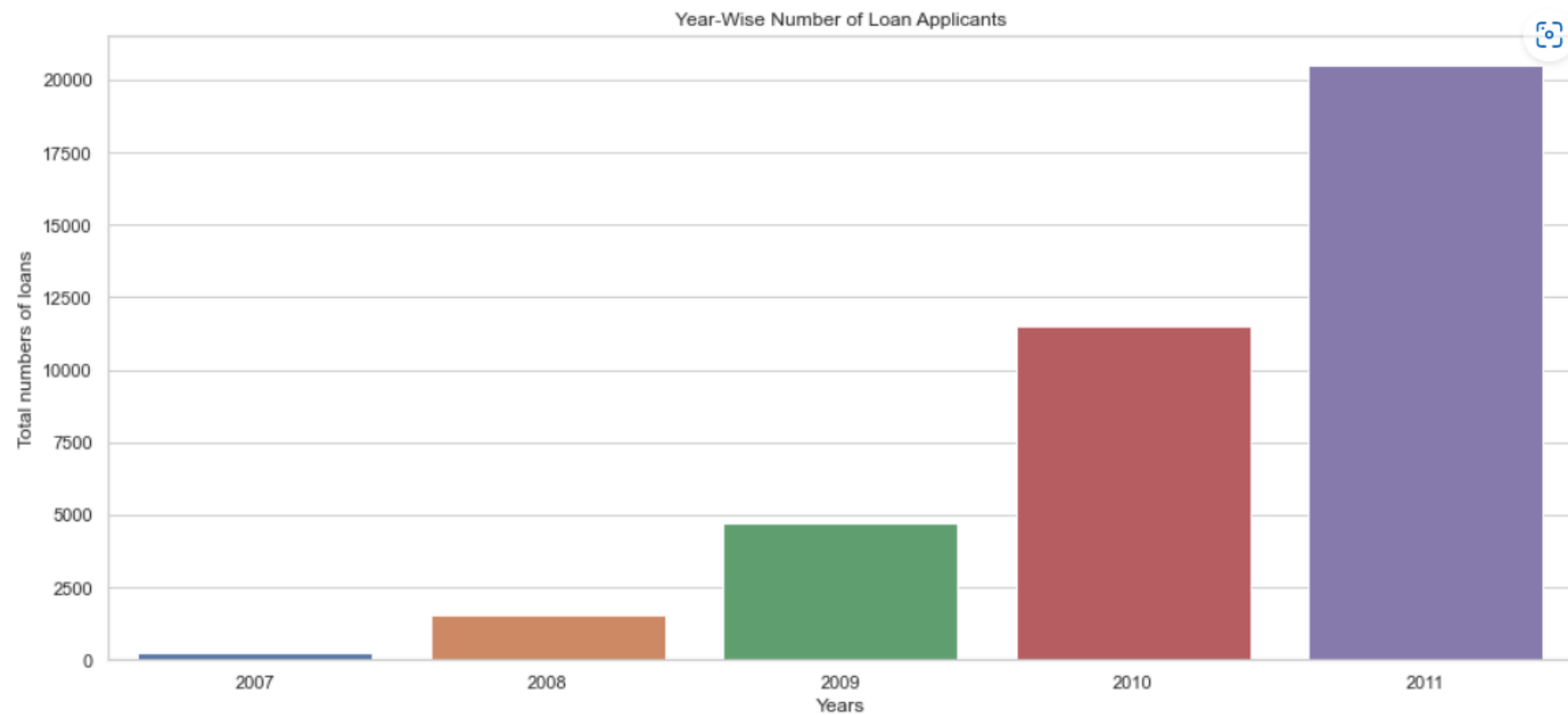
Customer's Demographic Information:

- emp_title,
- emp_length,
- home_ownership,
- annual_inc,
- verification_status,
- addr_state,
- zip_code,
- title,
- purpose,
- desc,
- url,

**Loan Characteristics Information:**

- loan_amnt,
- funded_amnt,
- funded_amnt_inv,
- int_rate,
- loan_status,
- grade,
- sub_grade,
- dti,
- loan_issue_d,
- term,
- installment,

Univariate Data

01
Univariate Data

This slide is 100% editable. Adapt it to your needs and capture your audience's attention.

02
Text Here

This slide is 100% editable. Adapt it to your needs and capture your audience's attention.

03
Text Here

This slide is 100% editable. Adapt it to your needs and capture your audience's attention.

Year-Wise Number of Loan Applicants

In this univariate analysis , Substantial growth between 2007-2011 in the number of loan applicants

Year-Wise Number of Loan Applicants

In this univariate analysis , Substantial growth between 2007-2011 in the number of loan applicants
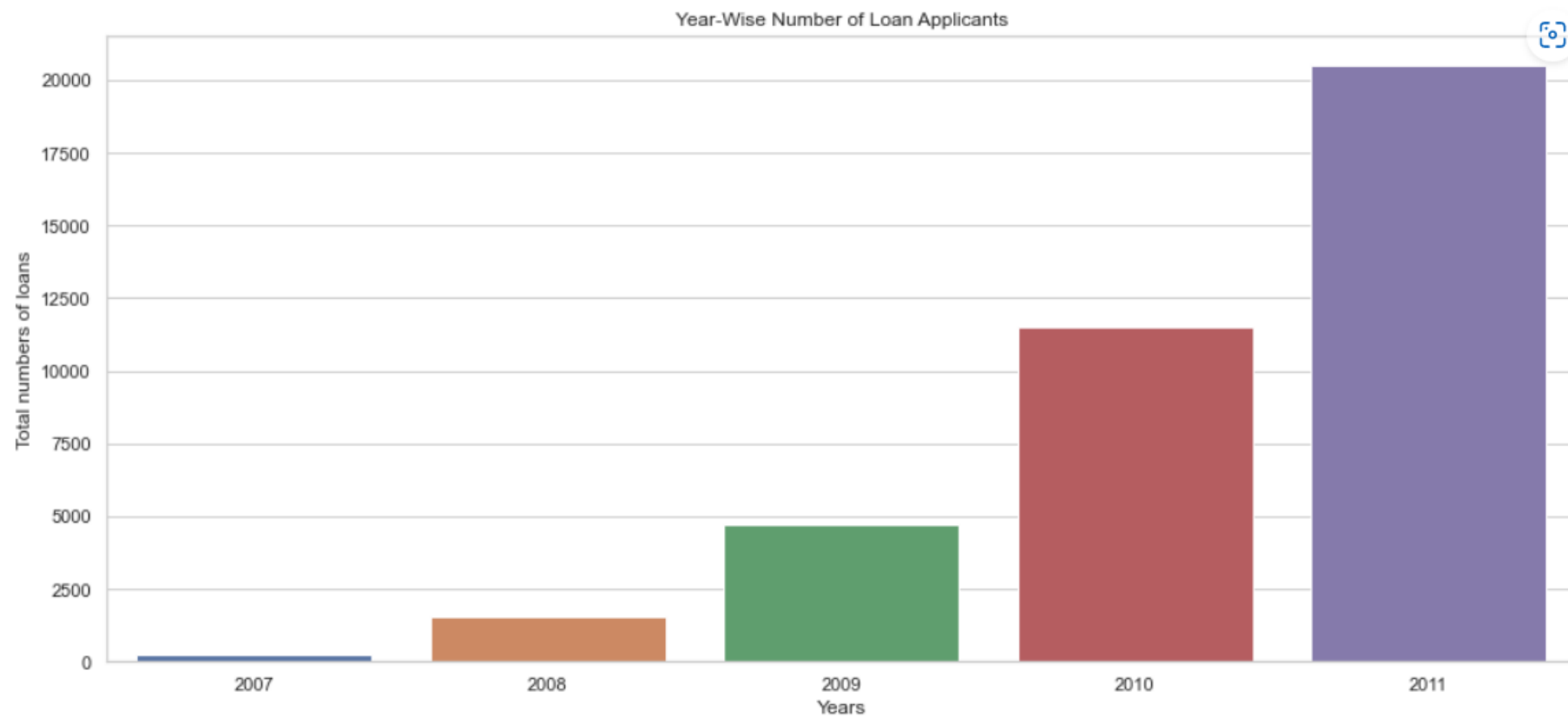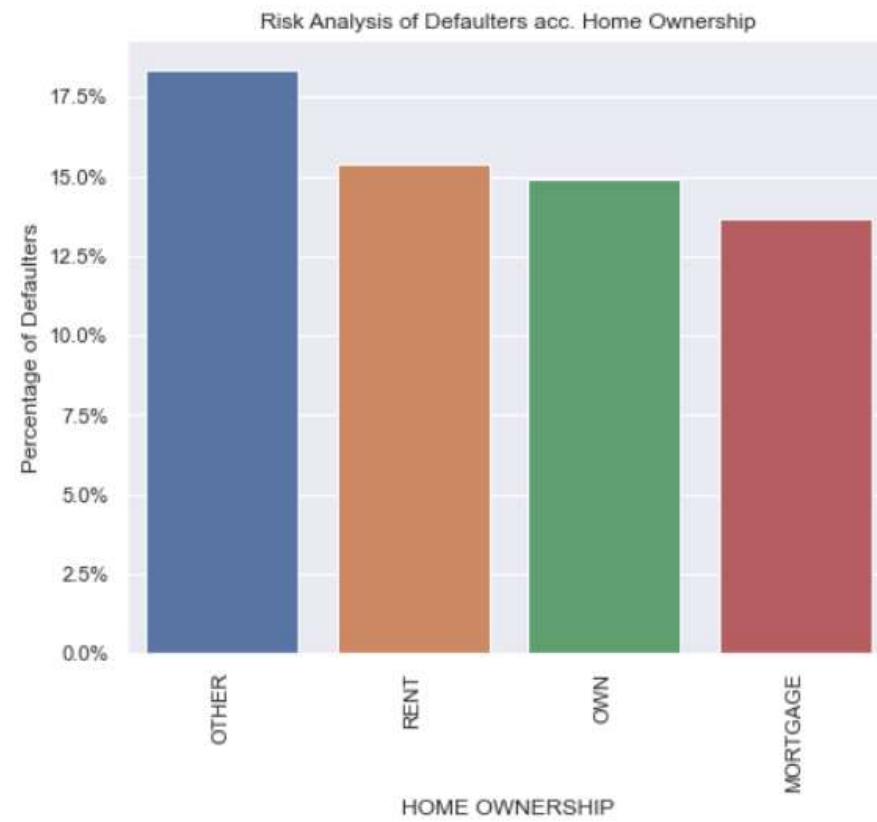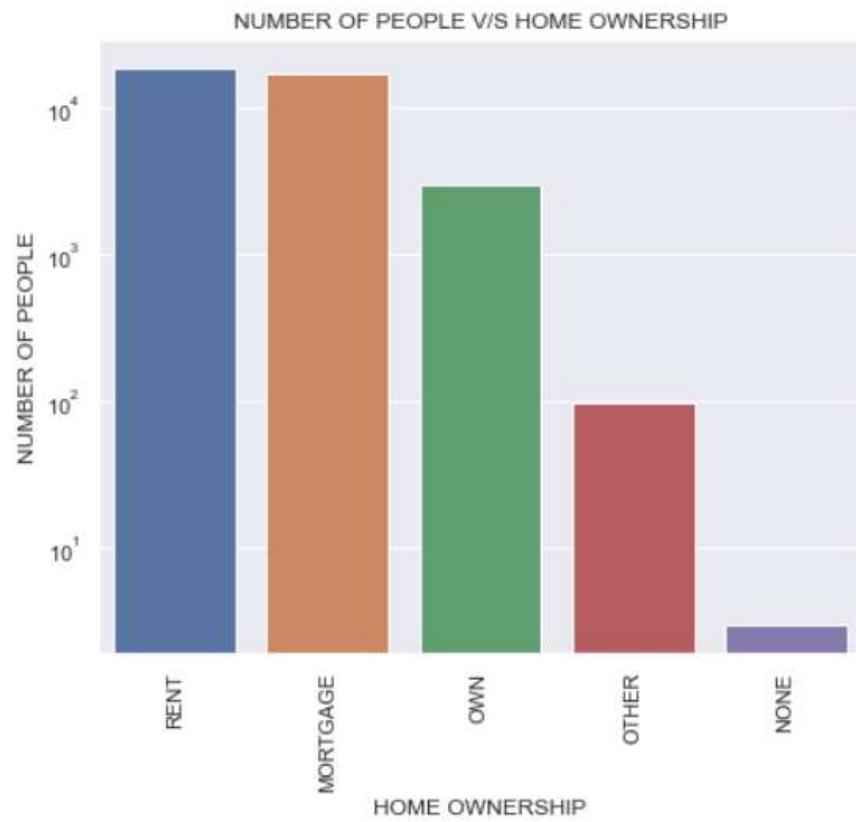
Distribution of loan amount
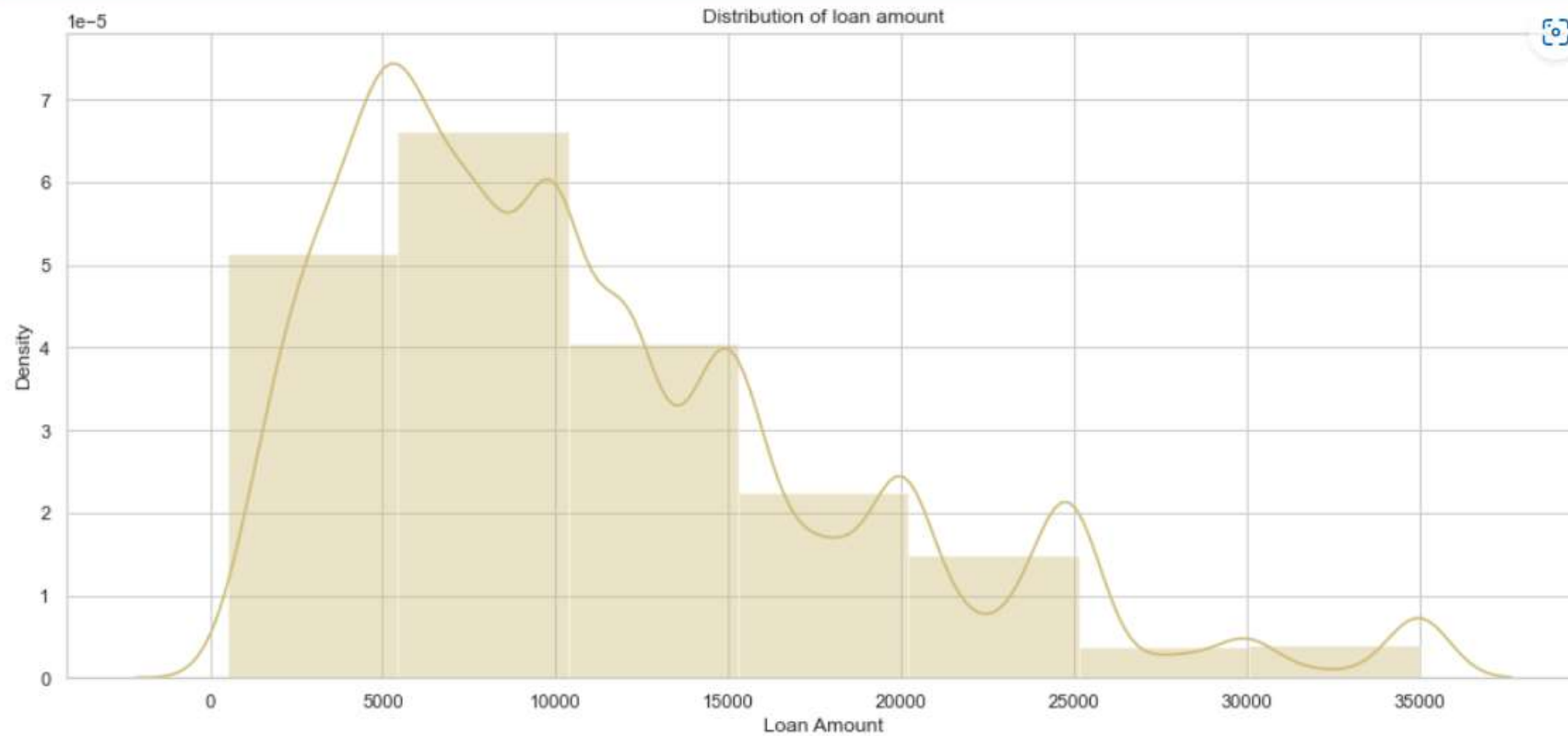
The distribution of loan amount i.e the amount of loan applied can be visualized from the above plot . It has been observed that the maximum loan range amount is between 5000-10000

Risk Analysis of Defaulters acc. to Grades
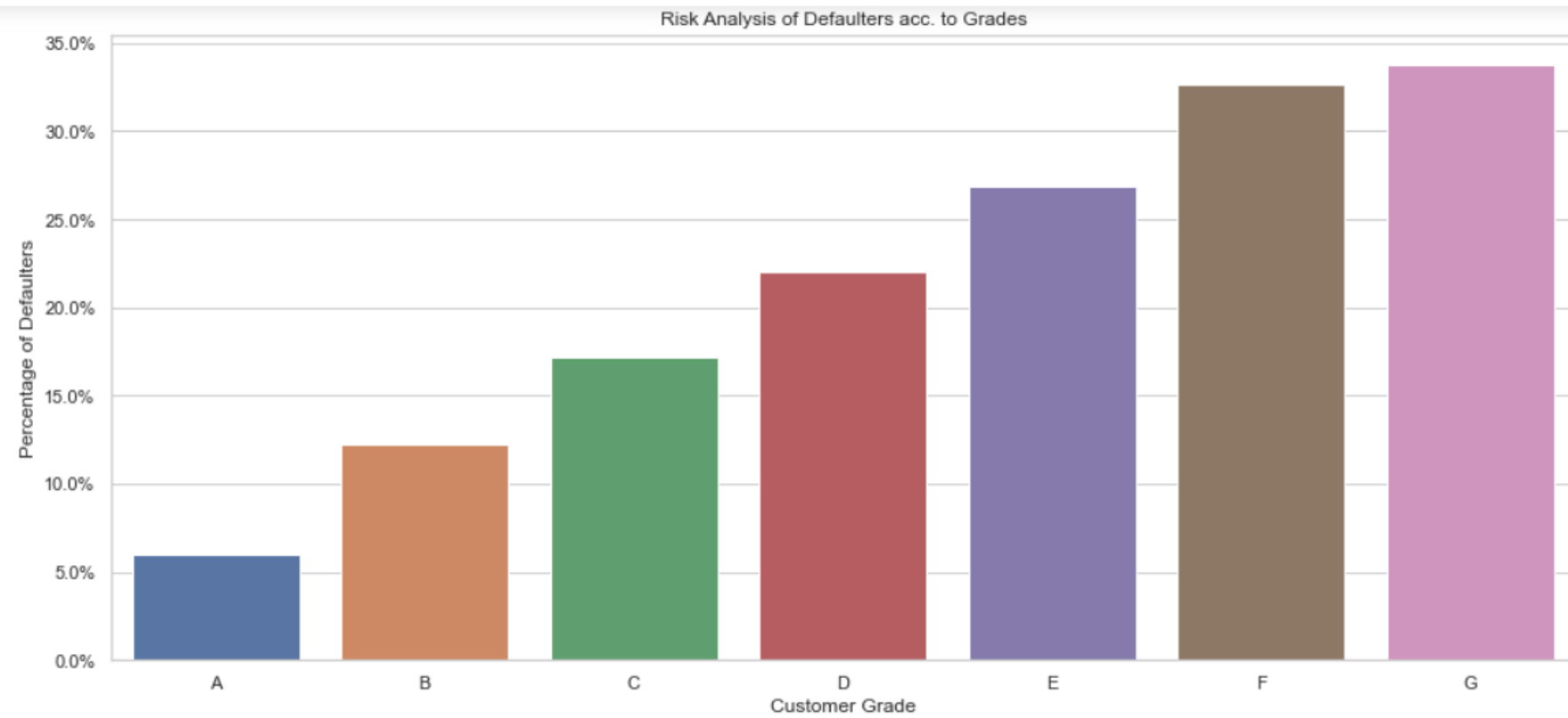
The percentage of Defaulters is the highest in Grade G group

Risk Analysis of Defaulters acc. to Purpose
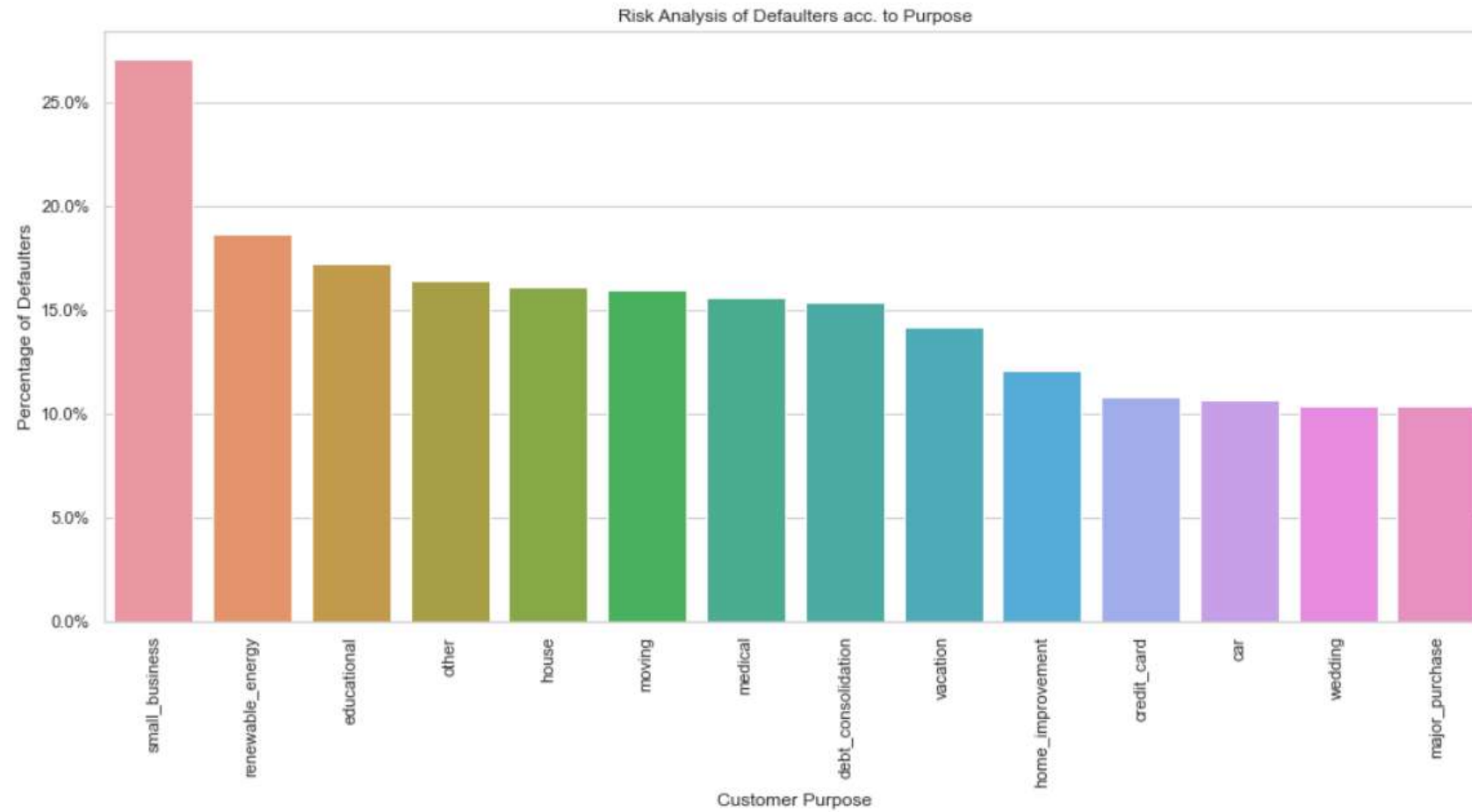
The percentage of risk of Defaulters is the largest in the Loan Purpose – 'small_business' , followed by 'renewable_energy' , 'educational' and so on.

NUMBER OF PEOPLE V/S HOME OWNERSHIP — Number of People vs Home Ownership (RENT, MORTGAGE, OWN, OTHER, NONE)

Risk Analysis of Defaulters acc. Home Ownership — Percentage of Defaulters vs Home Ownership (OTHER, RENT, OWN, MORTGAGE)

1. The number of people who have home ownership - "Rent"

2. The percentage of defaulter under home ownership is "Other" followed by "Rent" and "Own" and "Mortgage"
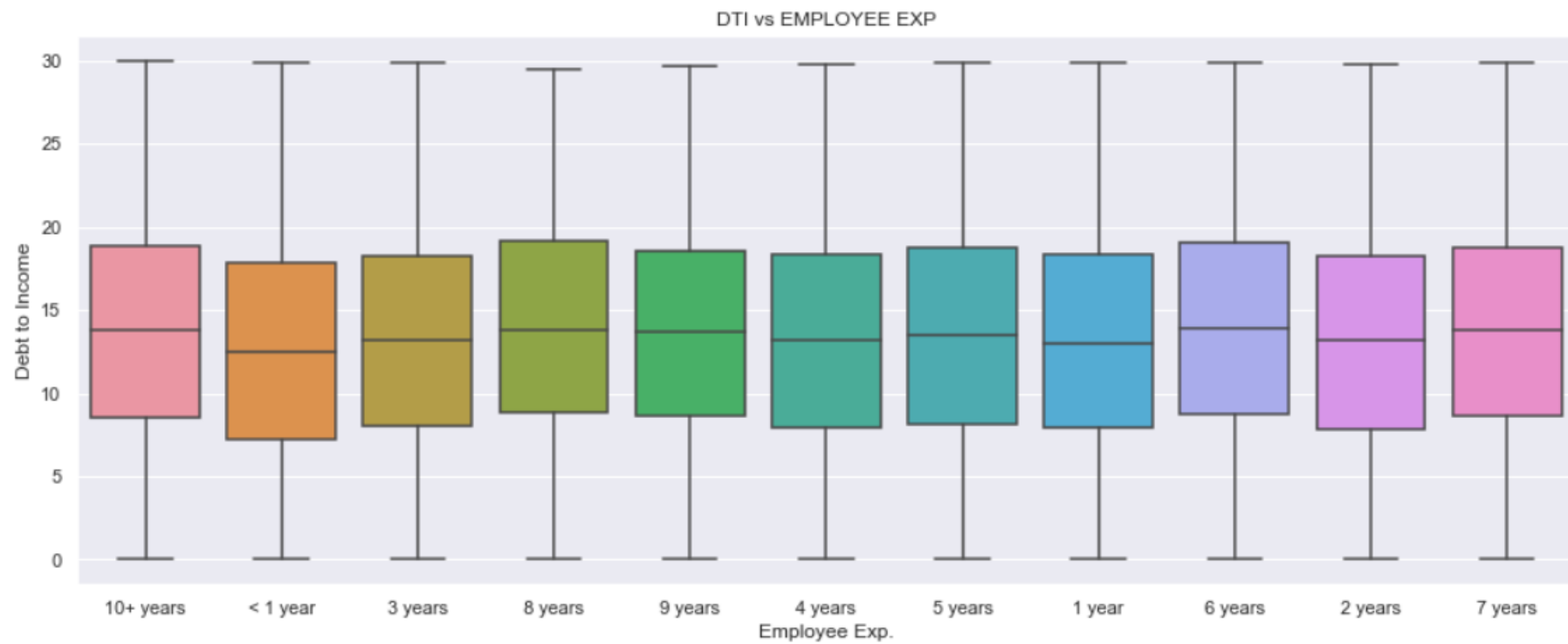
DTI vs EMPLOYEE EXP

By doing bivariate analysis it has been observed that 10+ years of employee exp. Is having debt to income in maximum range apart from that rest year of employee exp. are more or less in same range.

Loan Amount vs EMPLOYEE EXP
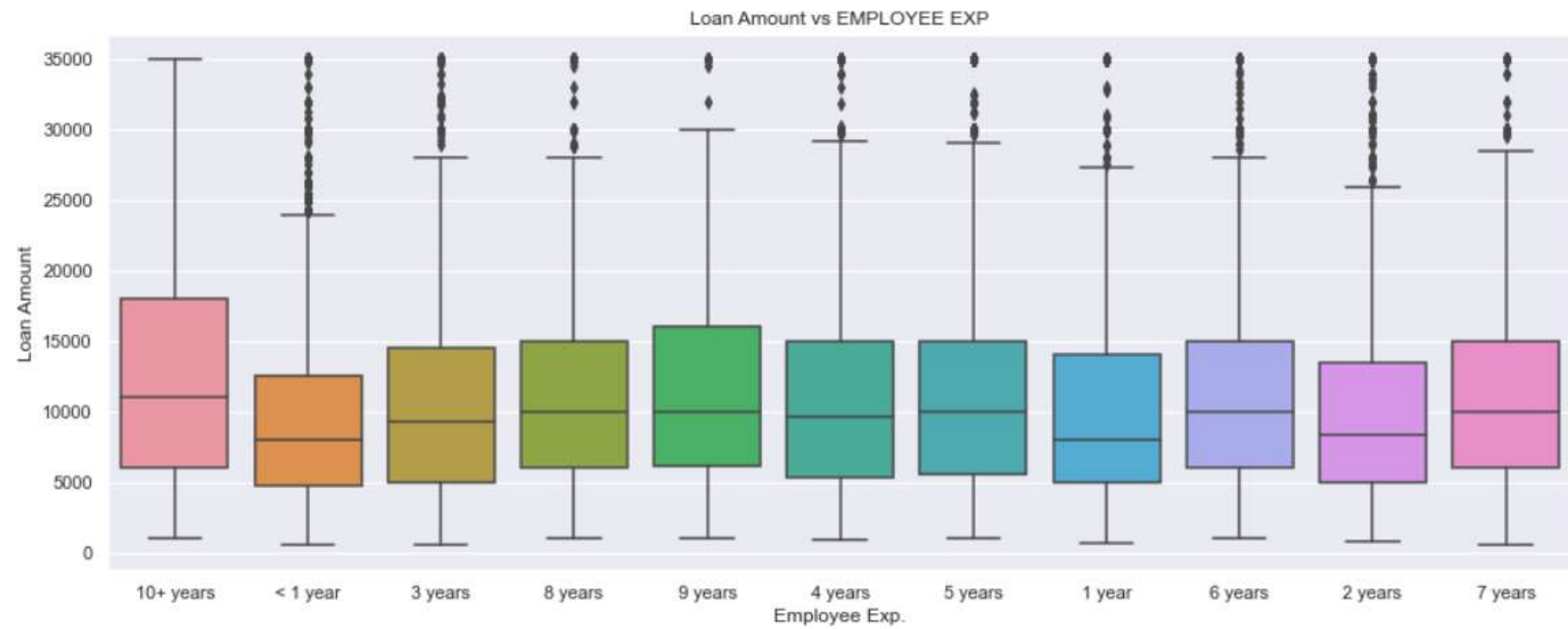
By doing bivariate analysis it has been observed that the employee exp. With 10+ years is taking more amount of loan and those who are less than 1 year is taken less amount of loan.

INTEREST RATE vs EMPLOYEE GRADES
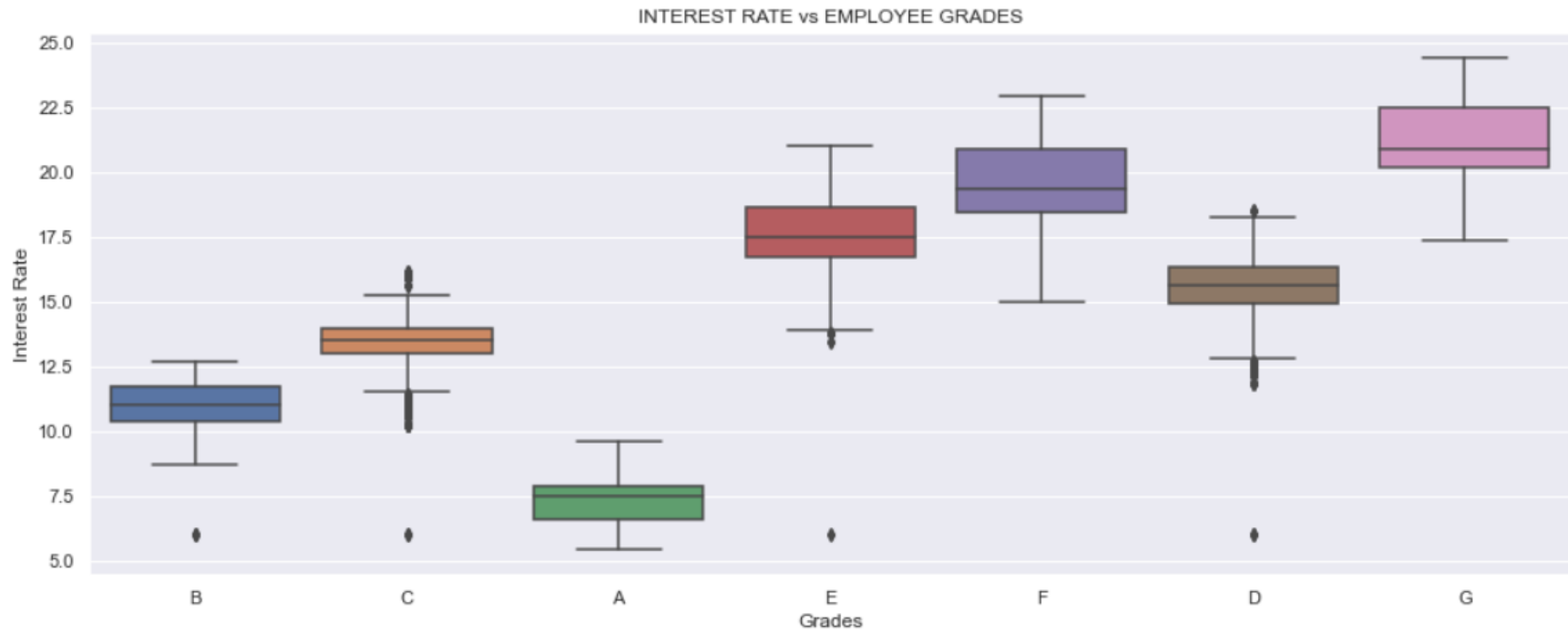
From the above plot it has been visualize that based on grade interest rate is going higher – "A" is having minimum interest rate, whereas "B" is having higher interest rate.

Risk Analysis of Defaulters acc. Employee State (TOP 10)
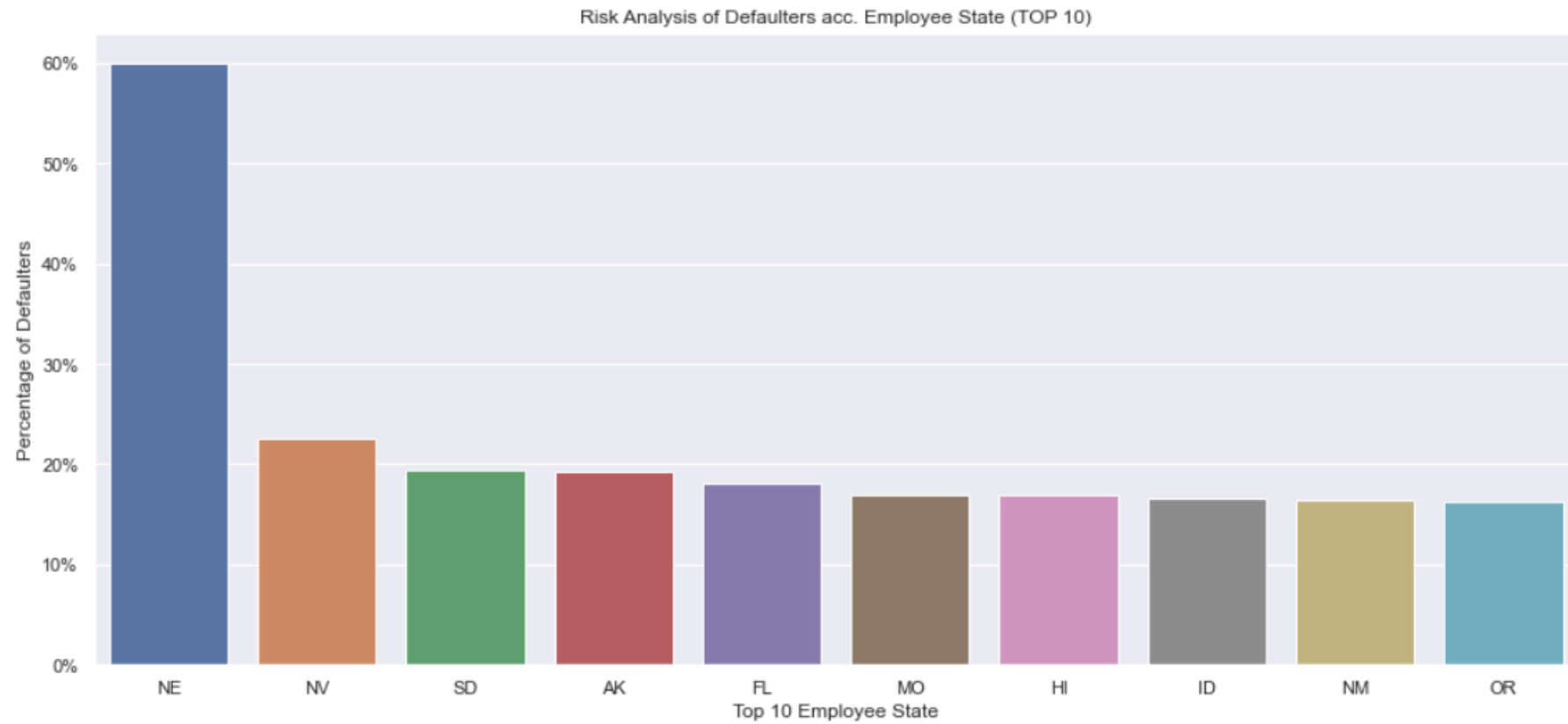
The percentage of Defaulters is maximum from NE (address) which is 60%

Conclusion:

Risk Analysis of Defaulters acc. Loan Term)