

EDA: LOAN Assessment Case Study Project

Submitted By :
Akash Sikroria



Problem Statement

Introduction

Solving this assignment will give you an idea about how real business problems are solved using EDA. In this case study, apart from applying the techniques you have learnt in EDA, you will also develop a basic understanding of risk analytics in banking and financial services and understand how data is used to minimise the risk of losing money while lending to customers.

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 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.

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.

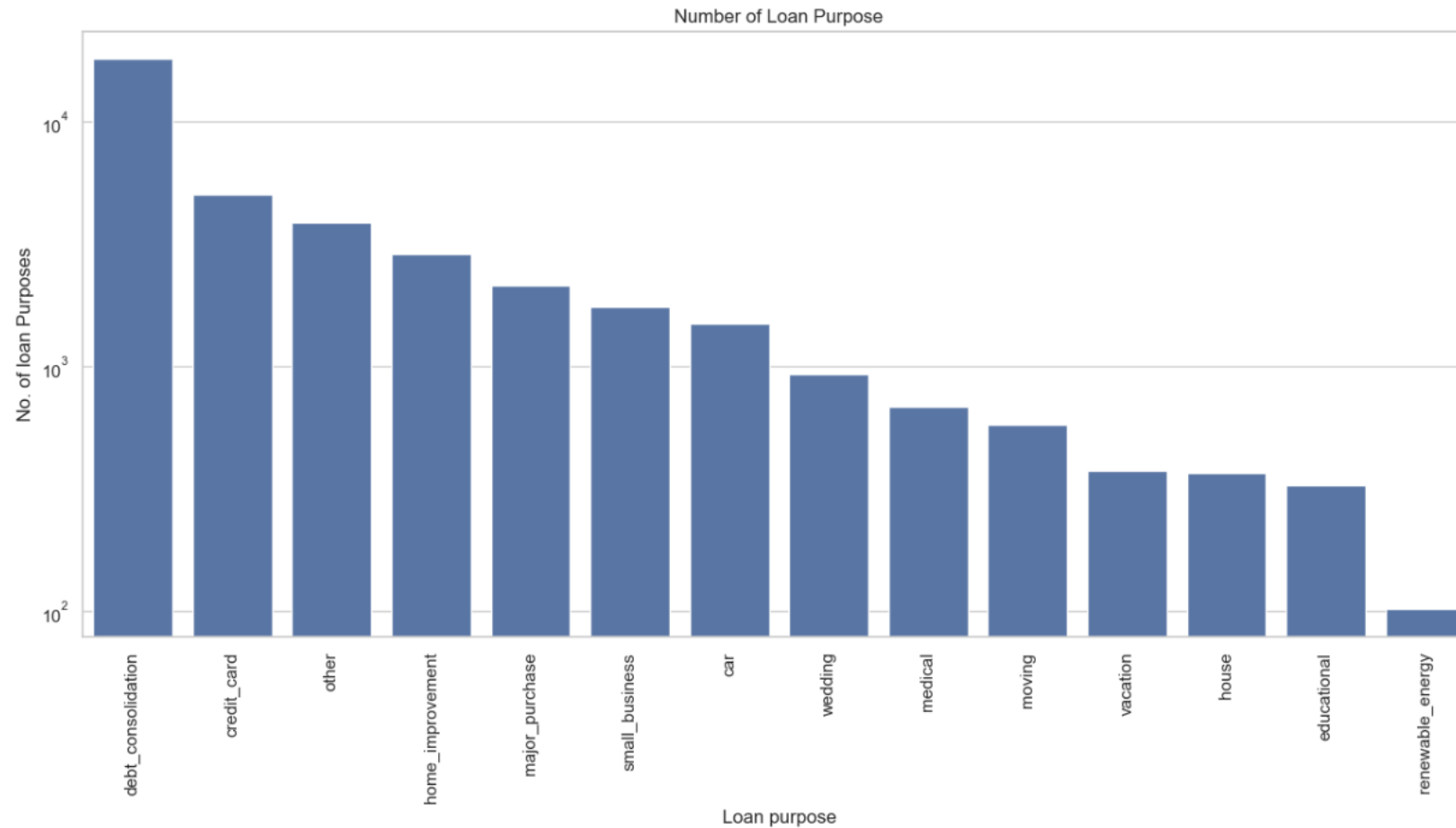
Like most other lending companies, lending loans to 'risky' applicants is the largest source of financial loss (called credit loss). Credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed. In other words, borrowers who **default** cause 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.

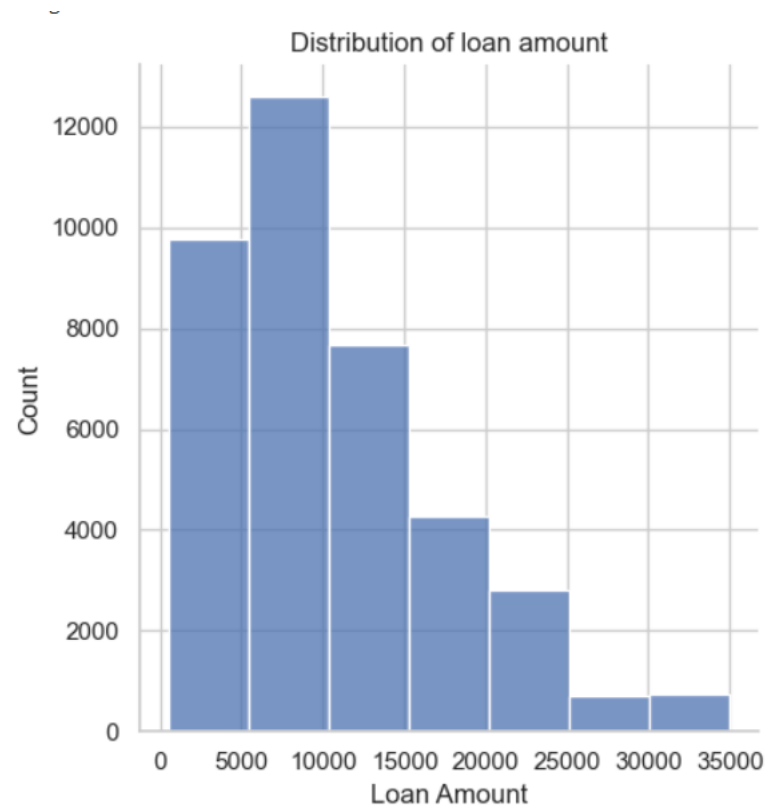
In other words, the company wants to understand the **driving factors (or driver variables)** behind loan default, i.e. the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.

To develop your understanding of the domain, you are advised to independently research a little about risk analytics (understanding the types of variables and their significance should be enough).

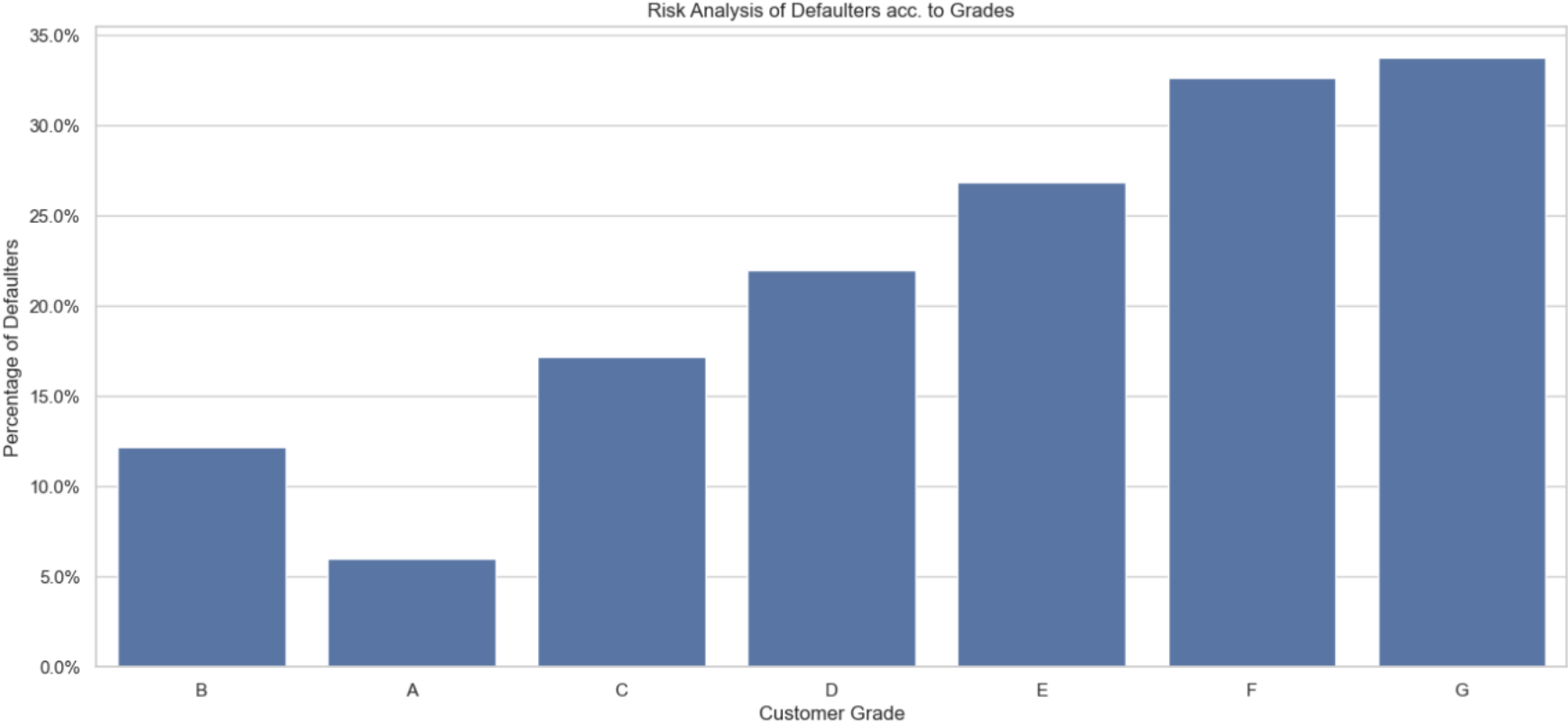
LOAN PURPOSE: The five major purposes of Loan applications are -
'debt_consolidation', 'credit_card', 'other', 'home_improvement' &
'major_purchase'
debt_consolidation' is the highest loan purpose



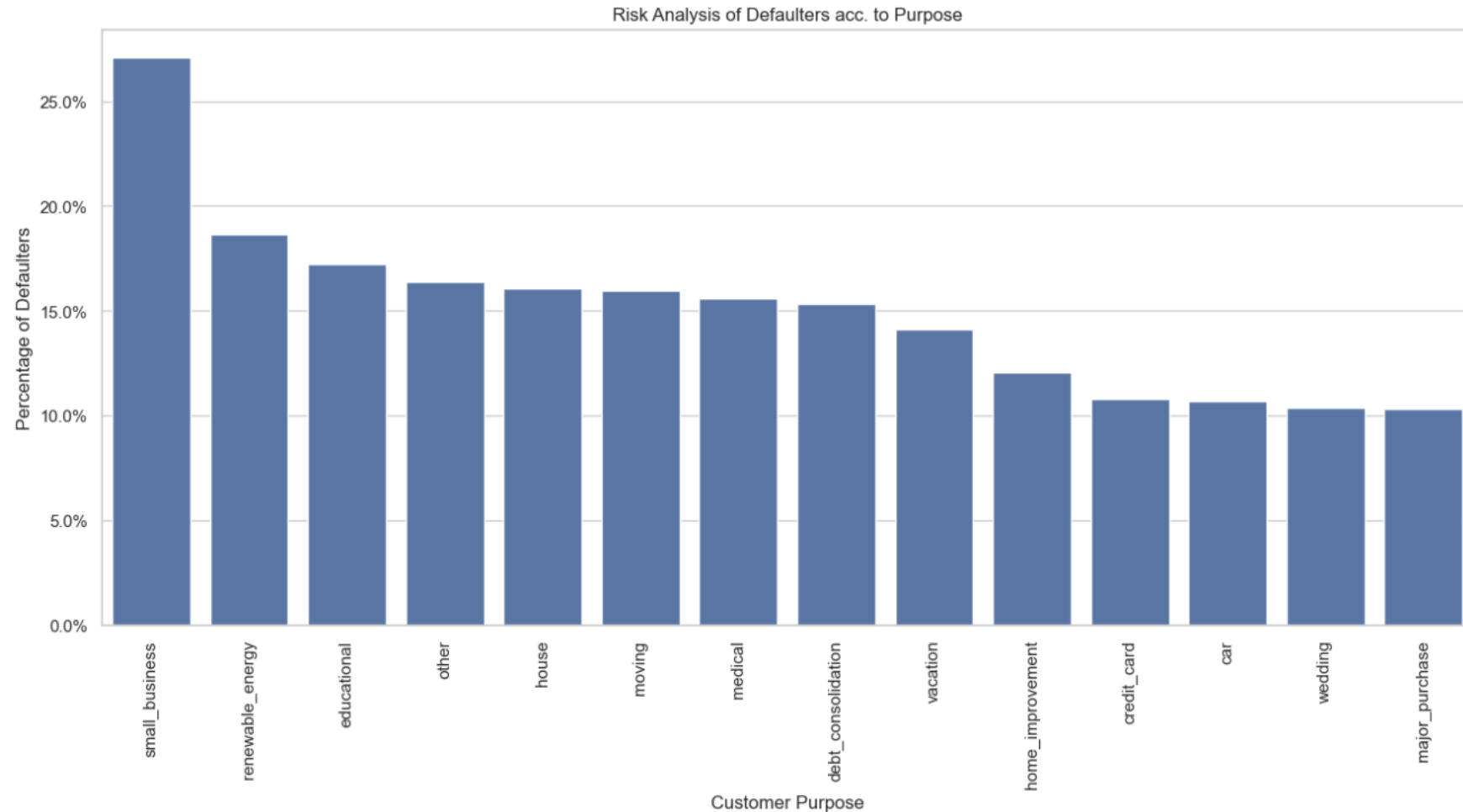
LOAN AMOUNT : The distribution of loan amount .It has been observed that the maximum loan range amount is between 5000-10000



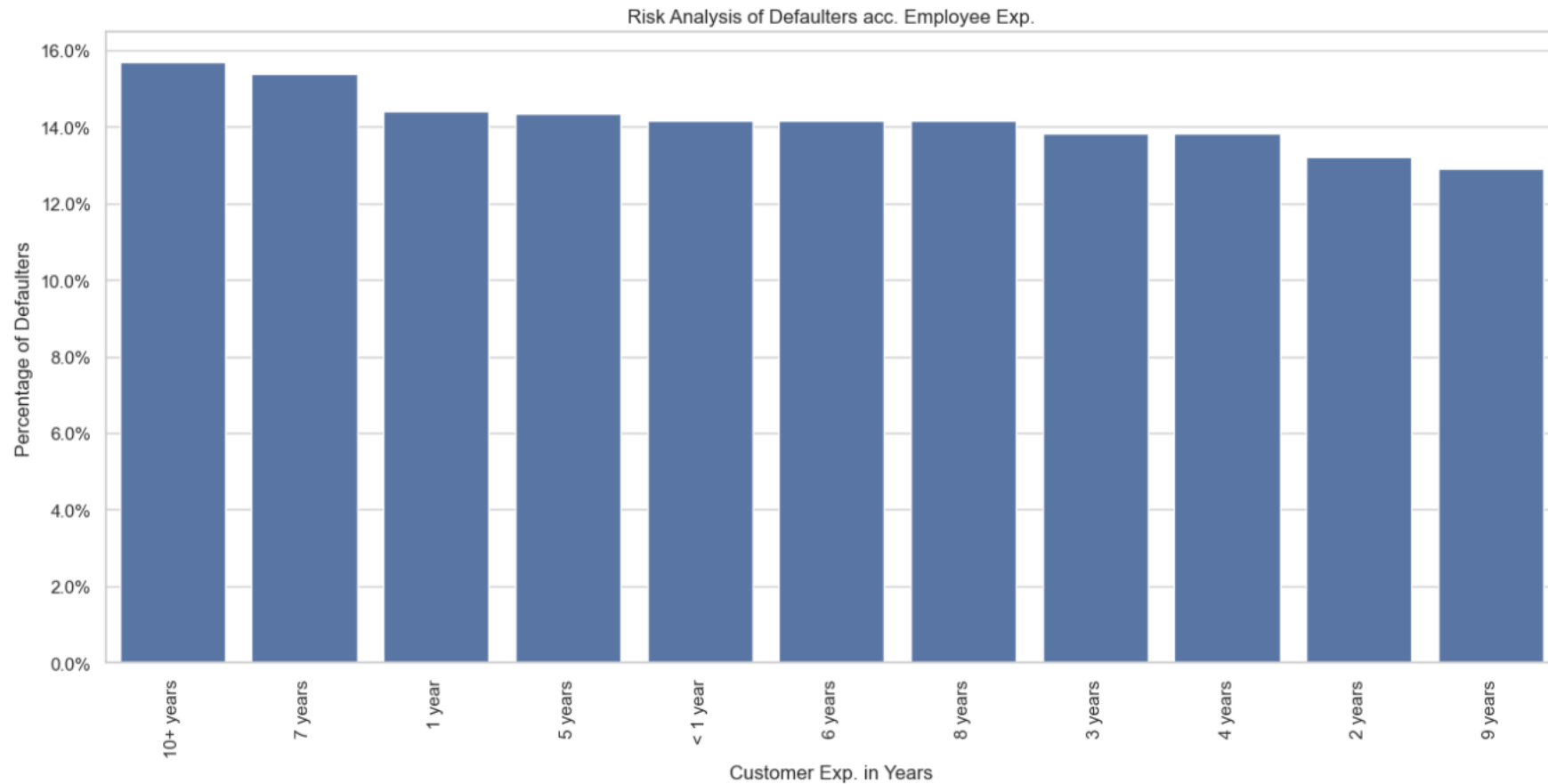
CUSTOMER GRADE: The percentage of Defaulters is the highest in Grade G group



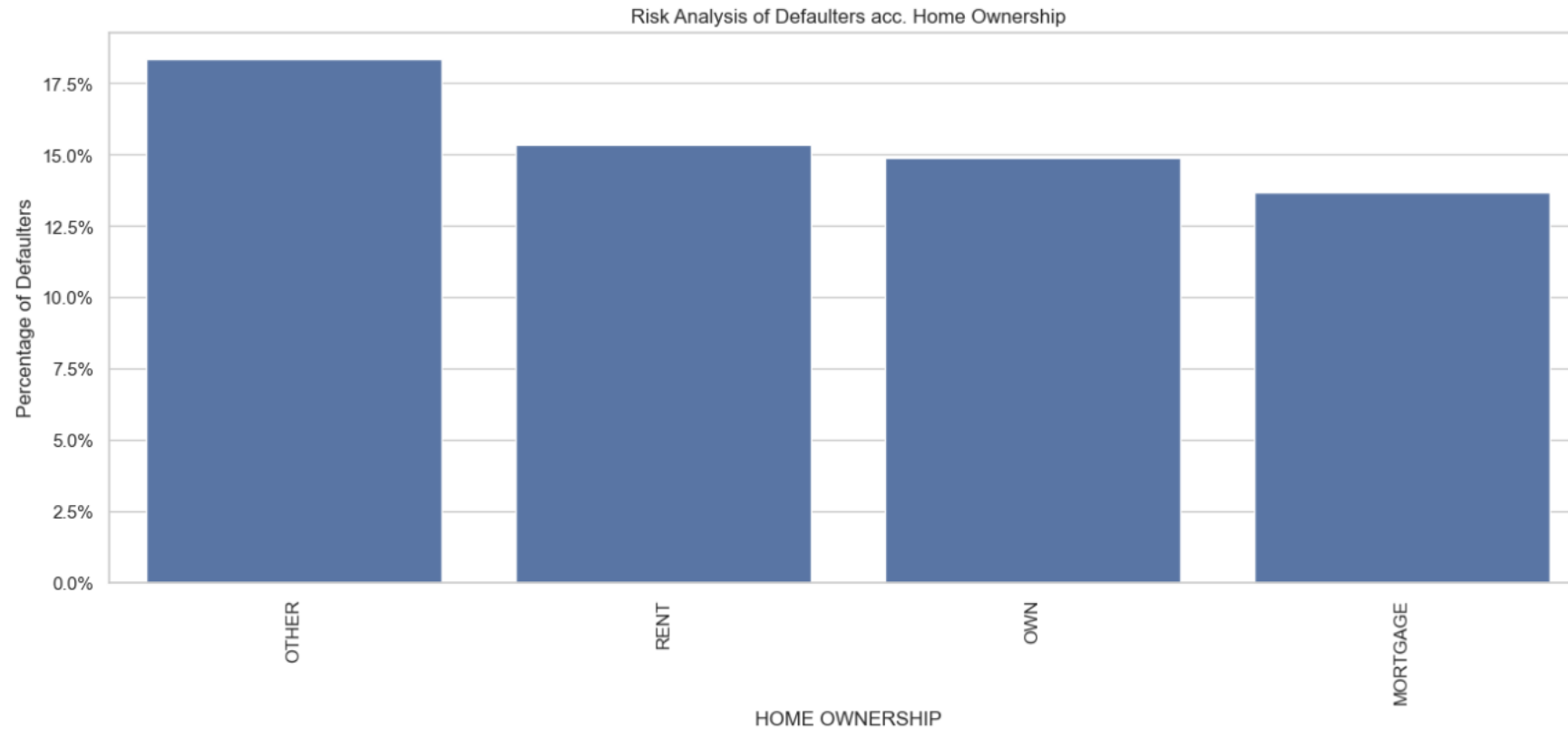
CUSTOMER PURPOSE: The percentage of risk of Defaulters is the largest in the Loan Purpose – ‘small_business’, followed by ‘renewable_energy’, ‘educational’



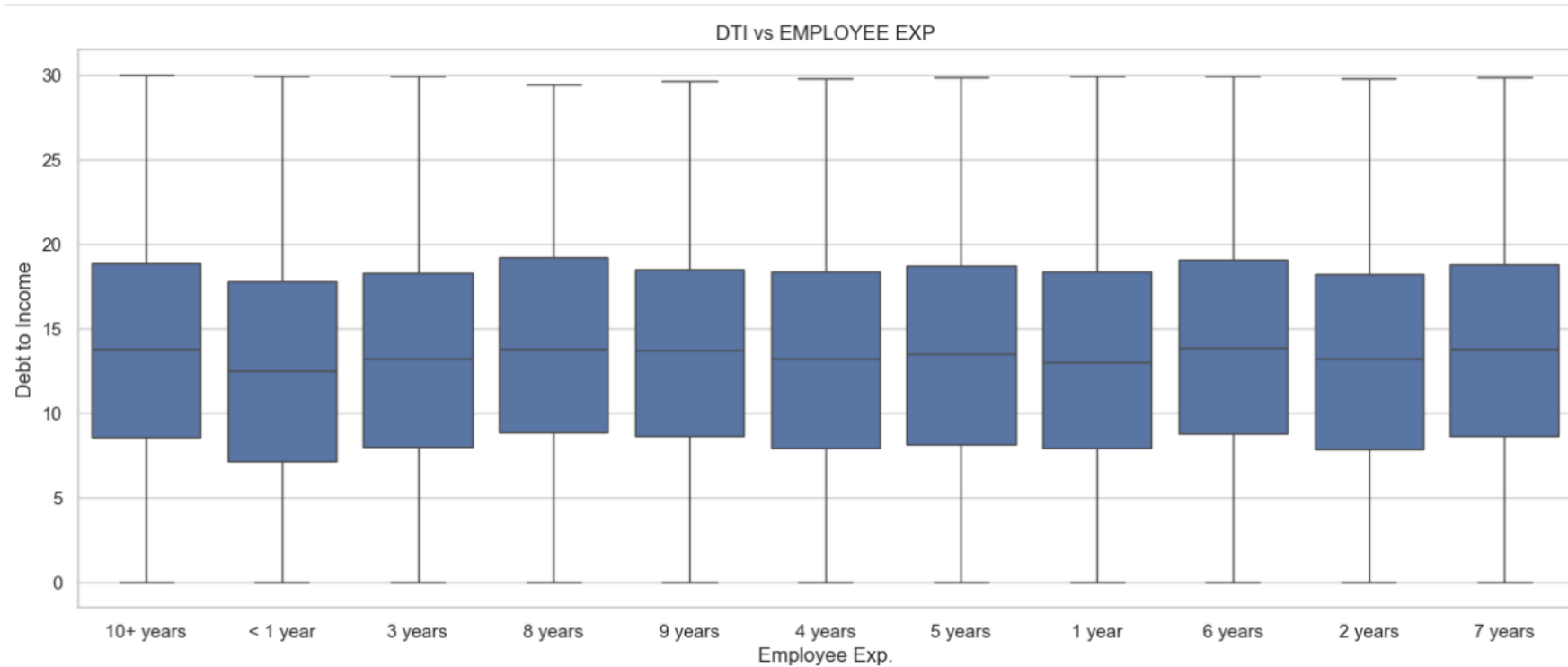
CUSTOMER EXPERIENCE: The percentage of risk of Defaulters is maximum for customer experience - 10+ years & minimum for 9 year's experience .



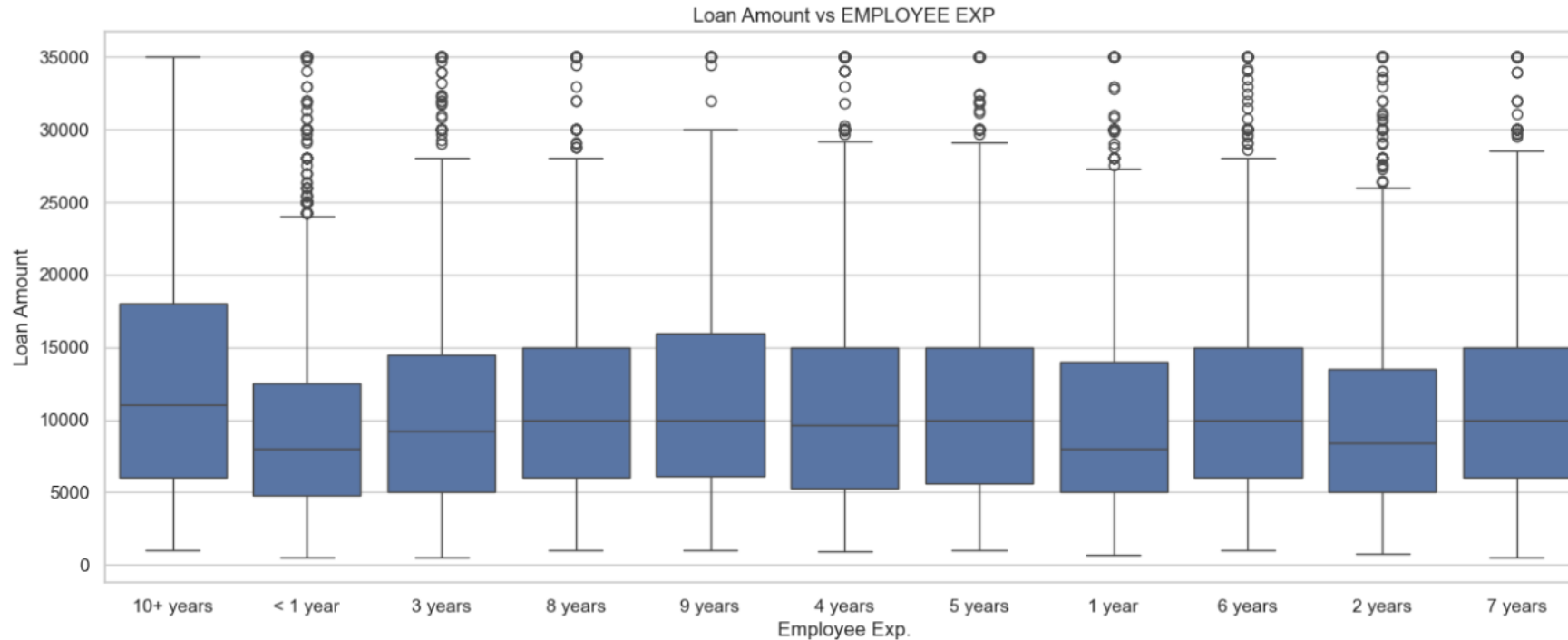
HOME OWNERSHIP: The percentage of defaulter under home ownership is “Other” followed by “Rent” and “Own” and “Mortgage”



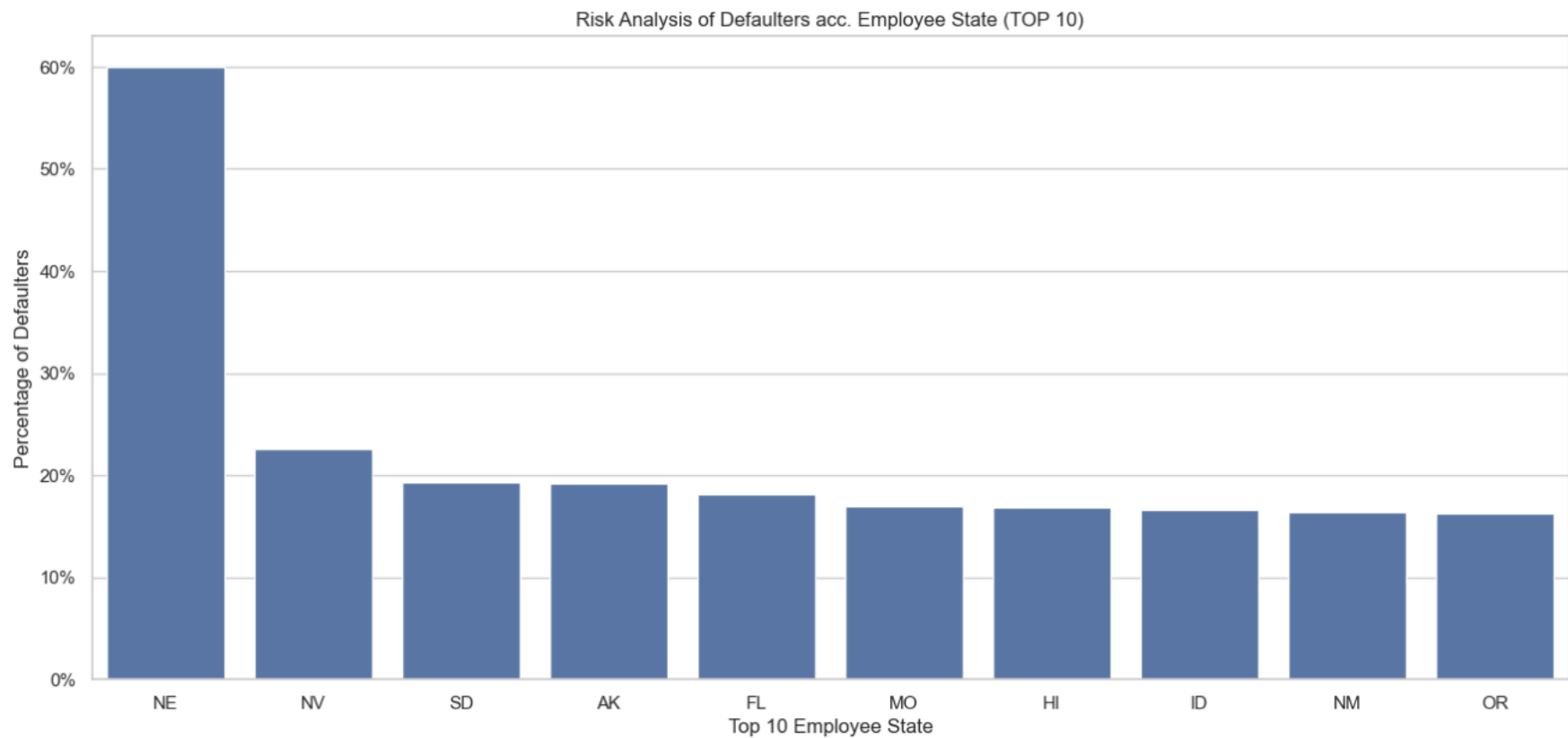
By doing bivariate analysis it has been observed that 10+ years of employee exp. Is having debt to income in maximum range apart from that rest year of employee exp. are more or less in same range.



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.



EMPLOYEE STATE: The percentage of Defaulters is maximum from NE (address) which is 60%



LOAN TERM: The percentage of Defaulters is found in 60 months as compared to 36 months – Loan Term

