LENDING CLUB CASE STUDY
BATCH: ML C36

AUTHORS:

1. SANGHAMITRA DEY
2. ABHIJIT GIRI

Contents

1. Introduction

We are given data about borrowers who have taken loans of various amounts from the Lending Club. The data involves various customer related information such as annual income, State, Loan Amount, Grade, Employment Length, Property type, Loan Purpose etc. Many other borrower performance related data is also present in the dataset. For each of the borrower we have been given if the borrower has been Fully Paid or has defaulted the loan.

2. Goal

The goal of this case study is to perform Exploratory Data Analysis on the various relevant data given to us and try to determine the driving factors which can tell if a borrower is likely to default the loan.

Data cleaning and Manipulation Approach

Below are the approach and methodology used to proceed with EDA and derive the insights:

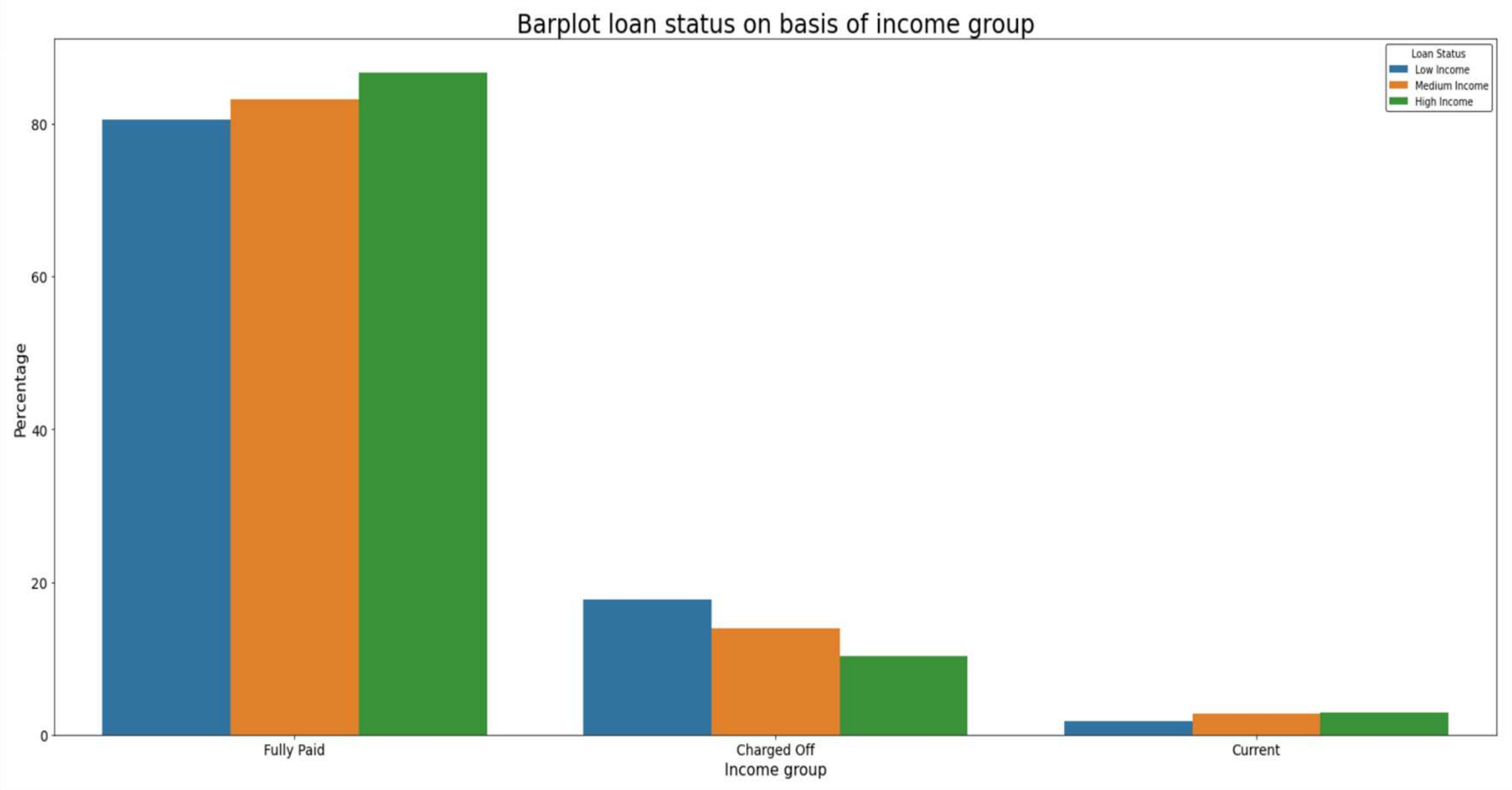
1. Import the data into Pandas from the given .csv file. The supplied data dictionary is also used to get more information.
2. Removing columns having large amount of null values and convert the data into proper format.
3. Filter out columns which are not relevant for analysis and may not be used achieve the goal. Customer related performance variables are examples of such columns.
4. Analyze the variables and identify the outliers and remove them.
5. Creating additional categorial values using binning and derive additional business related variables from given data.
6. Following are the variables which will be analyzed and EDA will be performed on the same:
    - 6.1.Loan status
    - 6.2.Loan amount
    - 6.3.Annual income
    - 6.4.Grade
    - 6.5.Sub grade
    - 6.6.purpose
    - 6.7.Address state
    - 6.8.Interest rate
    - 6.9.Home ownership
7. Perform univariate , bivariate/multivariate analysis on driver variables and derive insights.

DATA ANALYSIS

1.Annual Income:

Conclusion drawn about Annual Income:

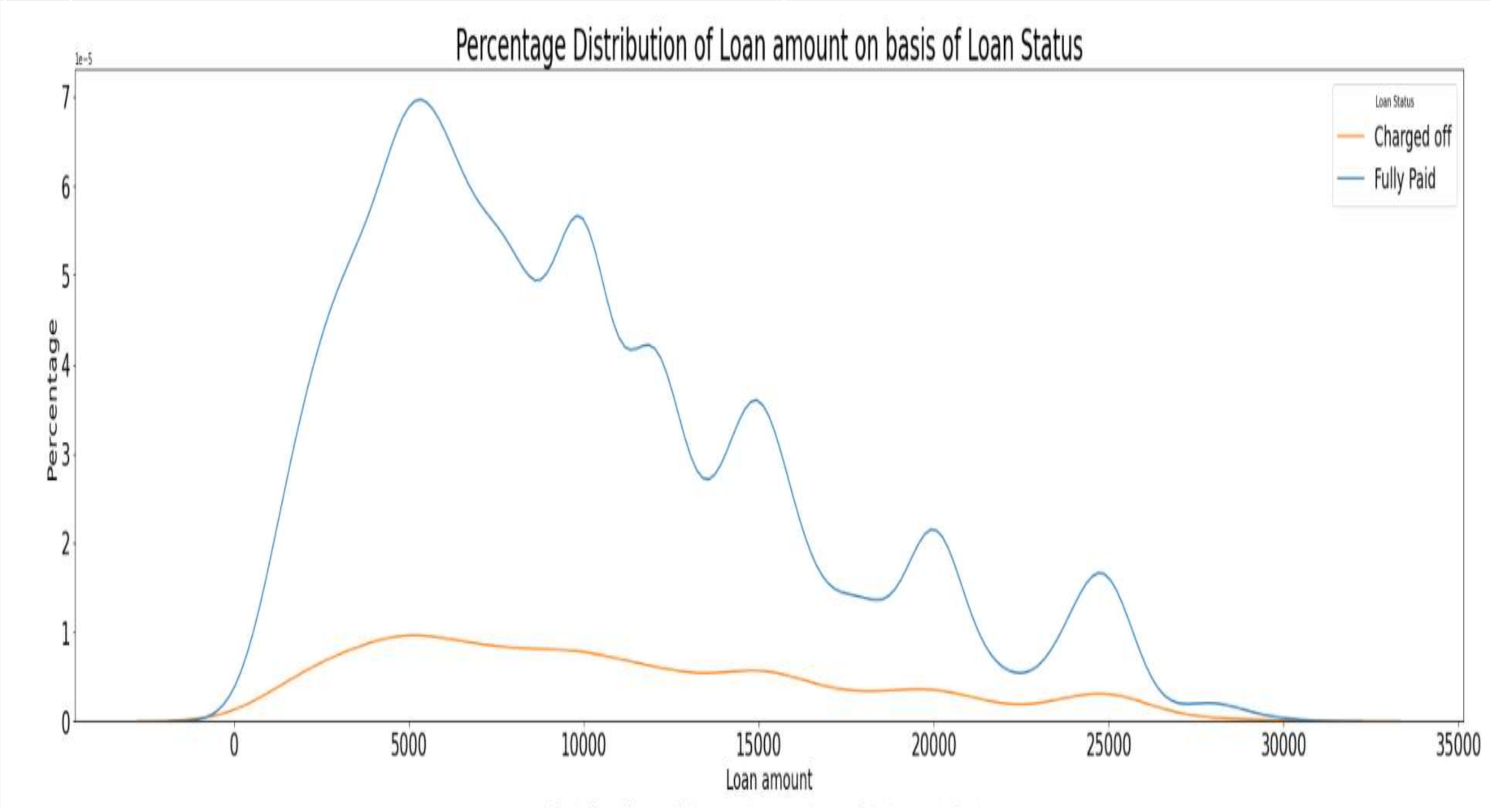
1. Borrowers who belong to low income category have higher chance of defaulting the loan.
2. The amount of loan taken by a borrower increases with the increase in annual income.
3. The chances of a borrower defaulting the loan is higher if his annual income is lower.



2.Loan Amount:

Conclusion drawn about Loan Amount:

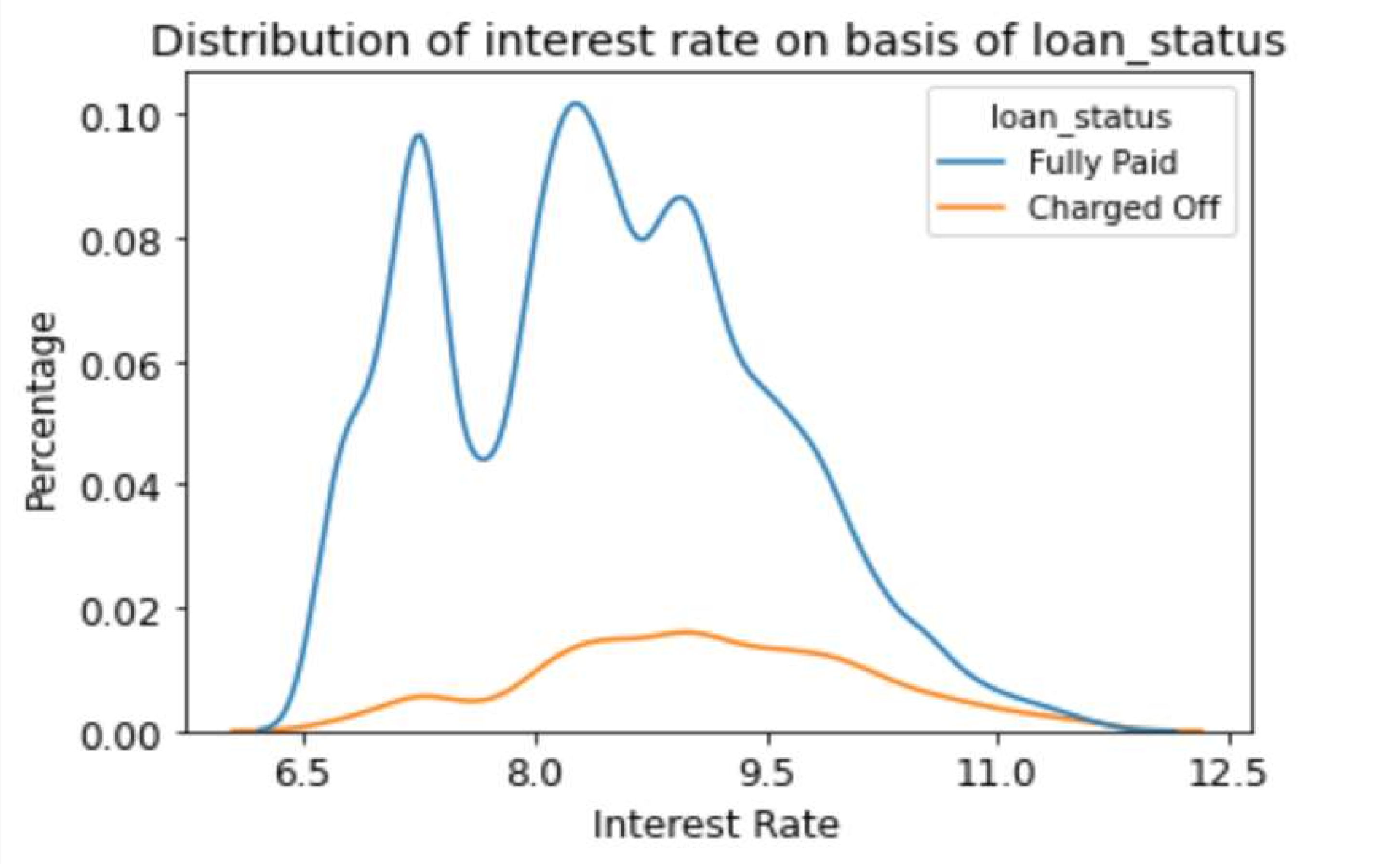
1.Most of the borrowers who have high chance of defaulting take loan amount between $5000 and $10000.



3.Interest Rate:

Conclusion drawn about Interest Rate:

1.For most borrowers who is charged interest rate between 8% to 9.5% have high chance of defaulting.

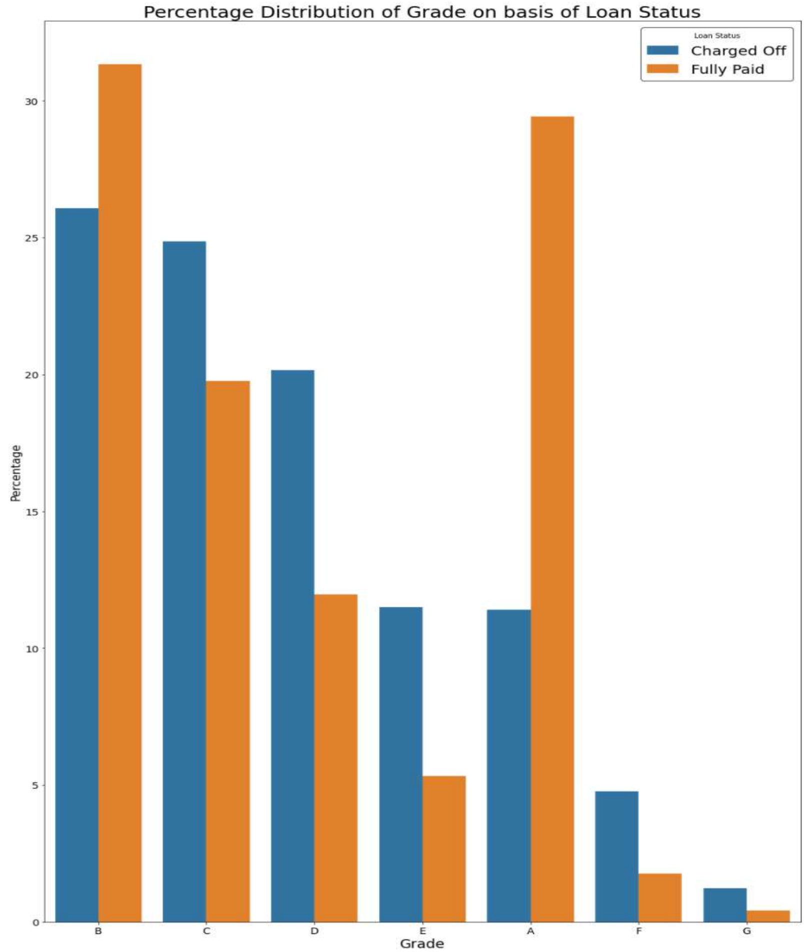


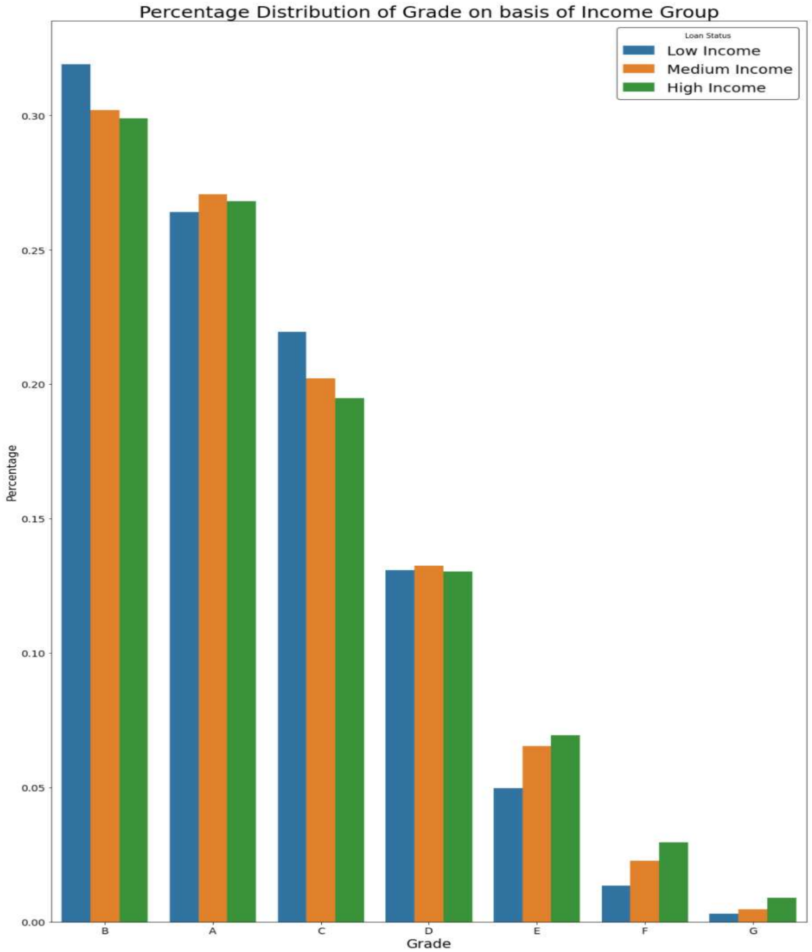
4.Grade:

Conclusion drawn about Grade:

1.Borrowers belonging to Grade C,D,E,F and G have higher chance of defaulting the loan irrespective of their income category.

2.Grade A and B people are less likely to default the loan.



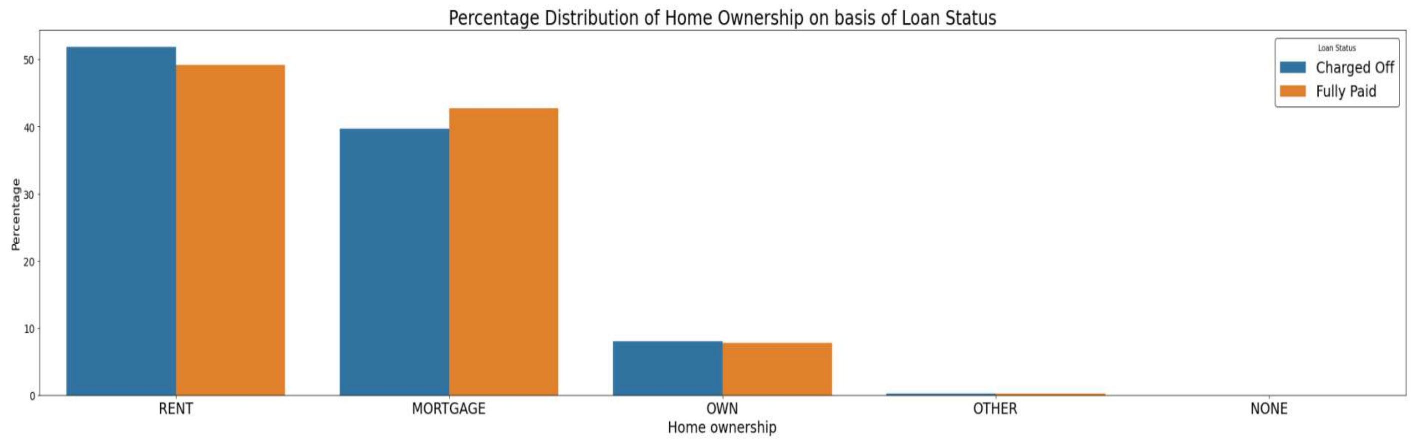


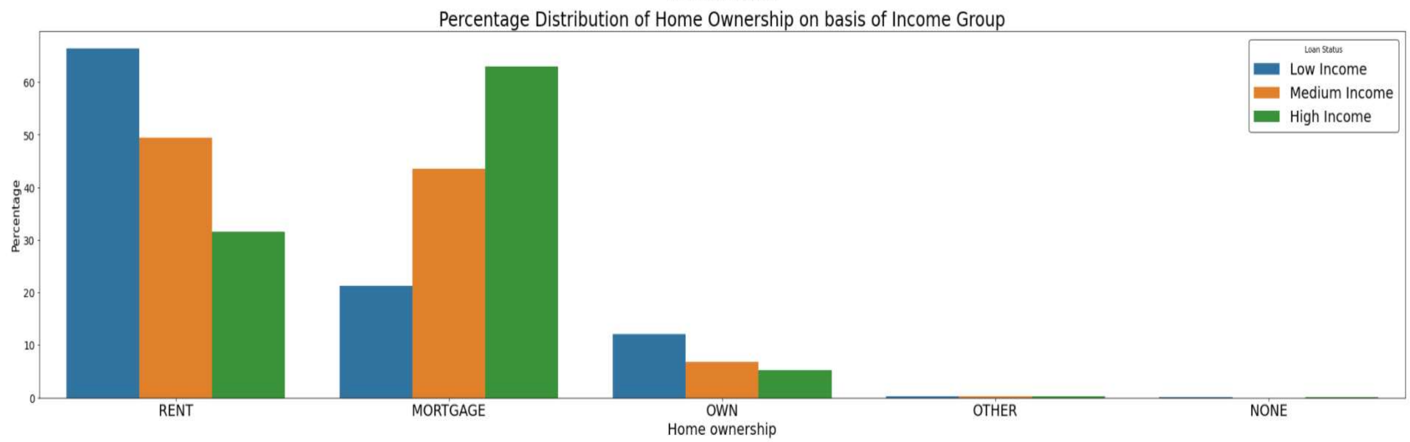
6.Home Ownership:

Conclusion drawn about Property ownership:

1.Borrowers living in Rented or mortgage property have greater chance of defaulting compared to borrowers having their own property.

2.Borrower belonging to low income category and staying in rented property have more chance of defaulting.

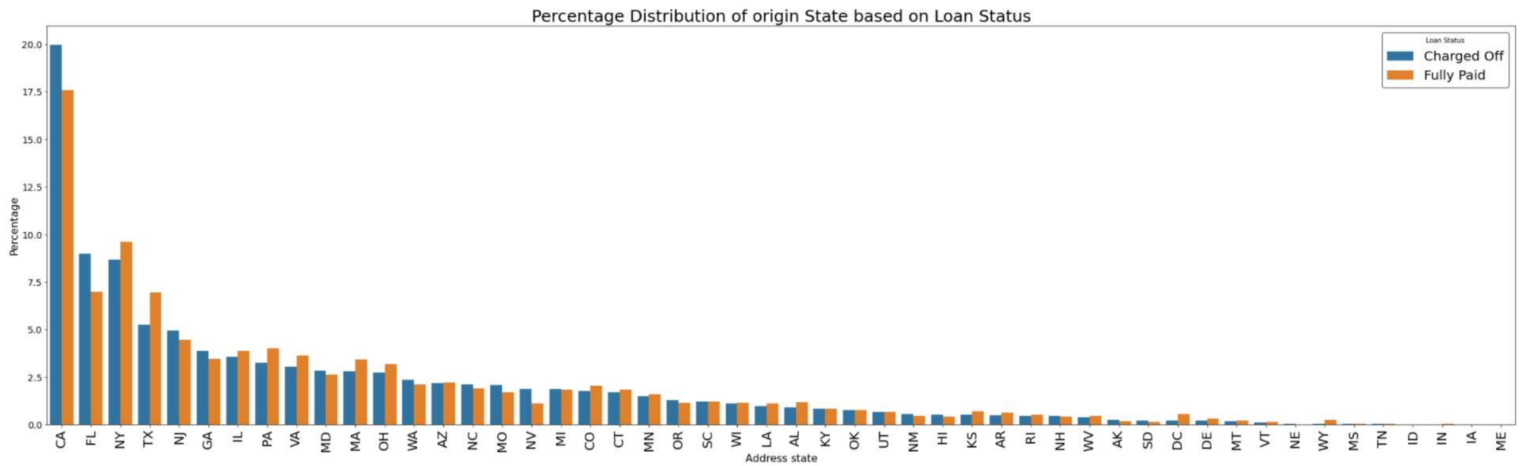


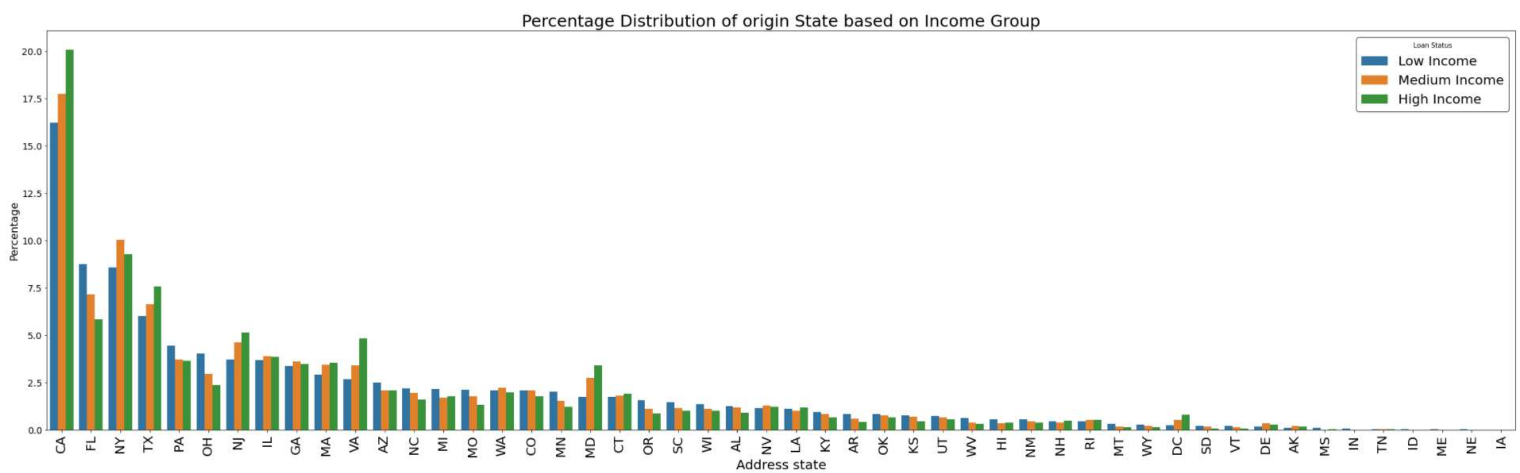


7.Address state:

Conclusion drawn about origin state:

1.Borrowers belonging to California(CA) ,New York(NY) and Florida(FL) have more chance of defaulting.
2.In most of the states people belonging to low income group have more chance of defaulting.



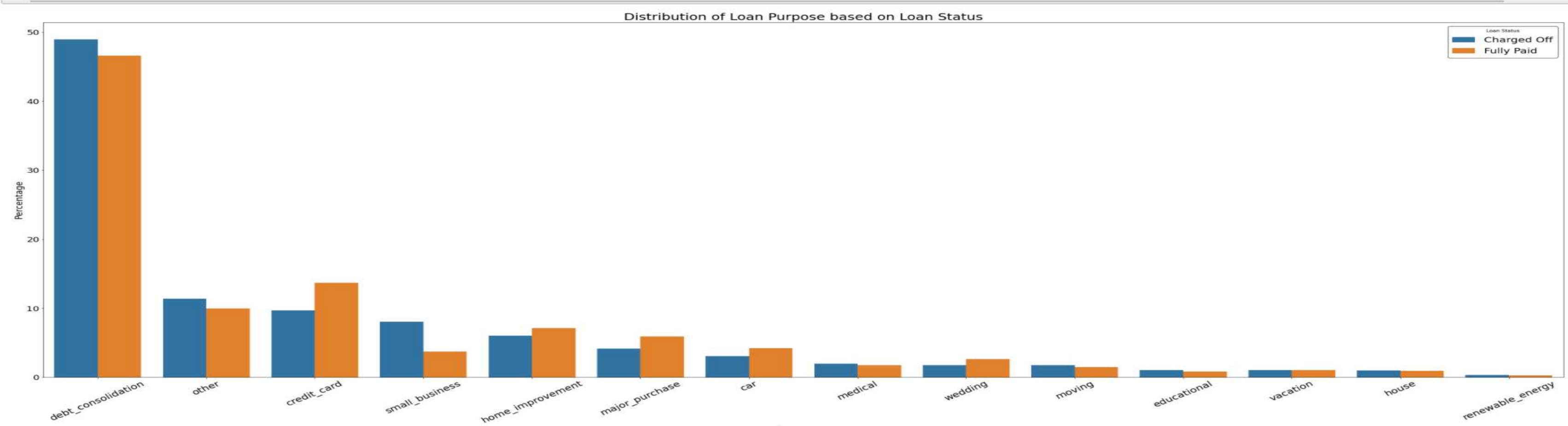


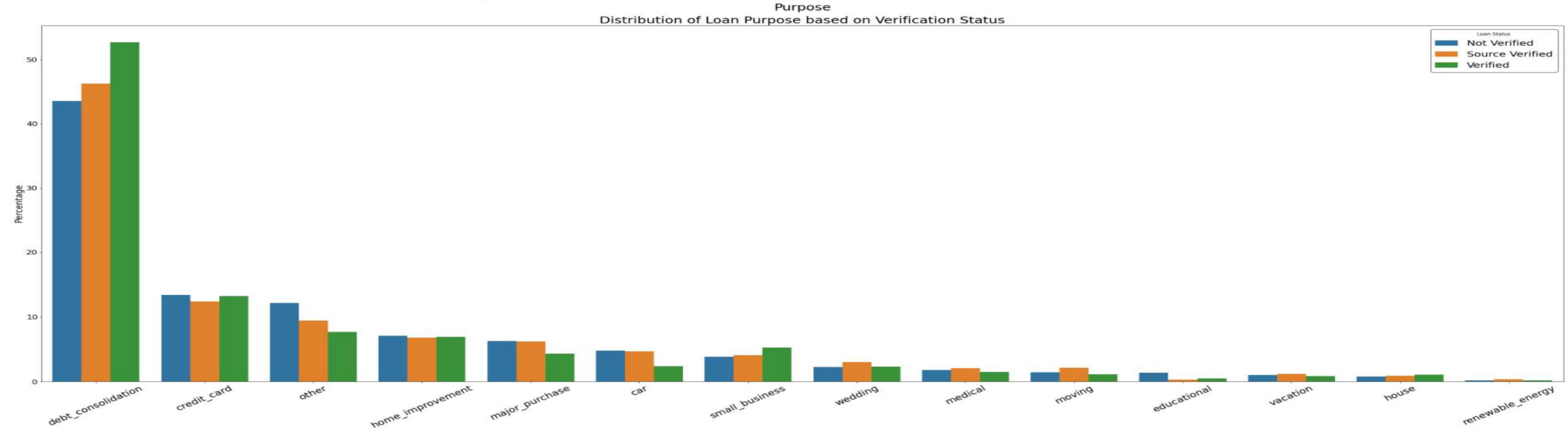
8.Loan Purpose:

Conclusion drawn about Loan Purpose

1.Borrowers taking loan for debt consolidation and opening small business have more chance of defaulting.

2.A Large number of unverified borrowers can be observed who are taking loan for "Other" purpose. These borrowers also have high chance of defaulting.

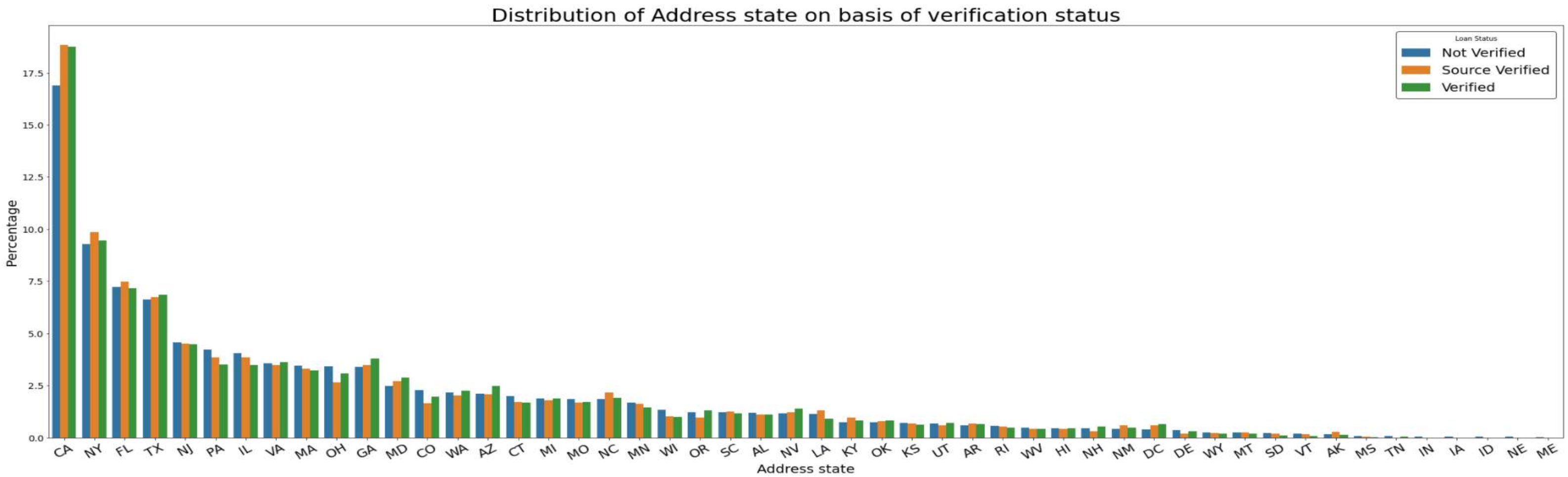


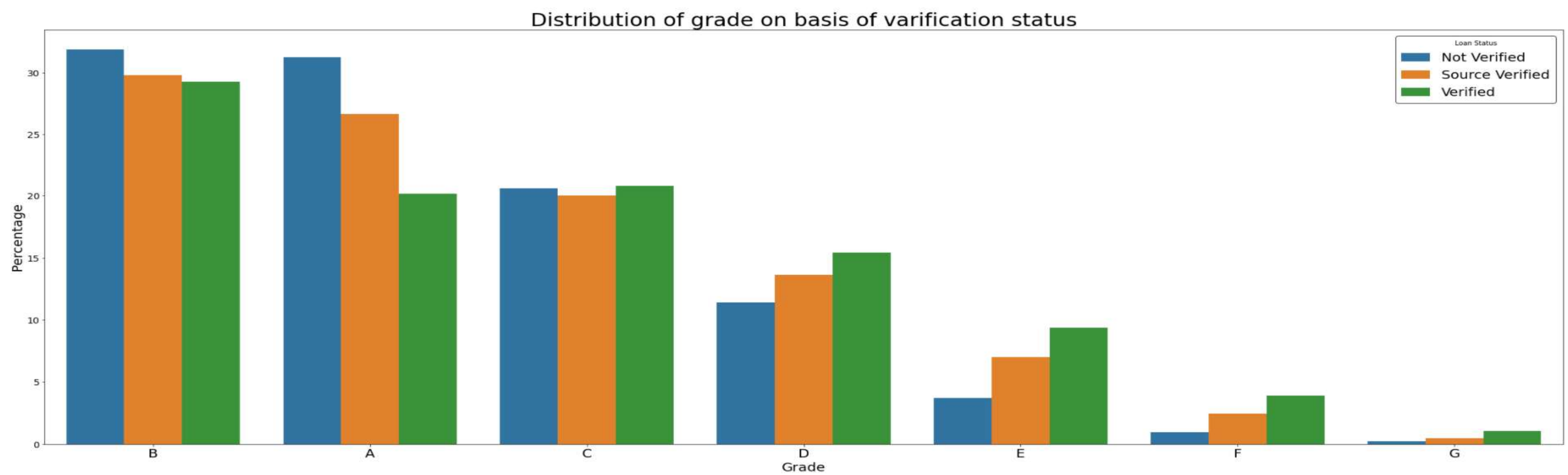


9.Verification status:

Conclusion drawn about Verification status

Verification status of person may not be able to predict if a person is likely to default.



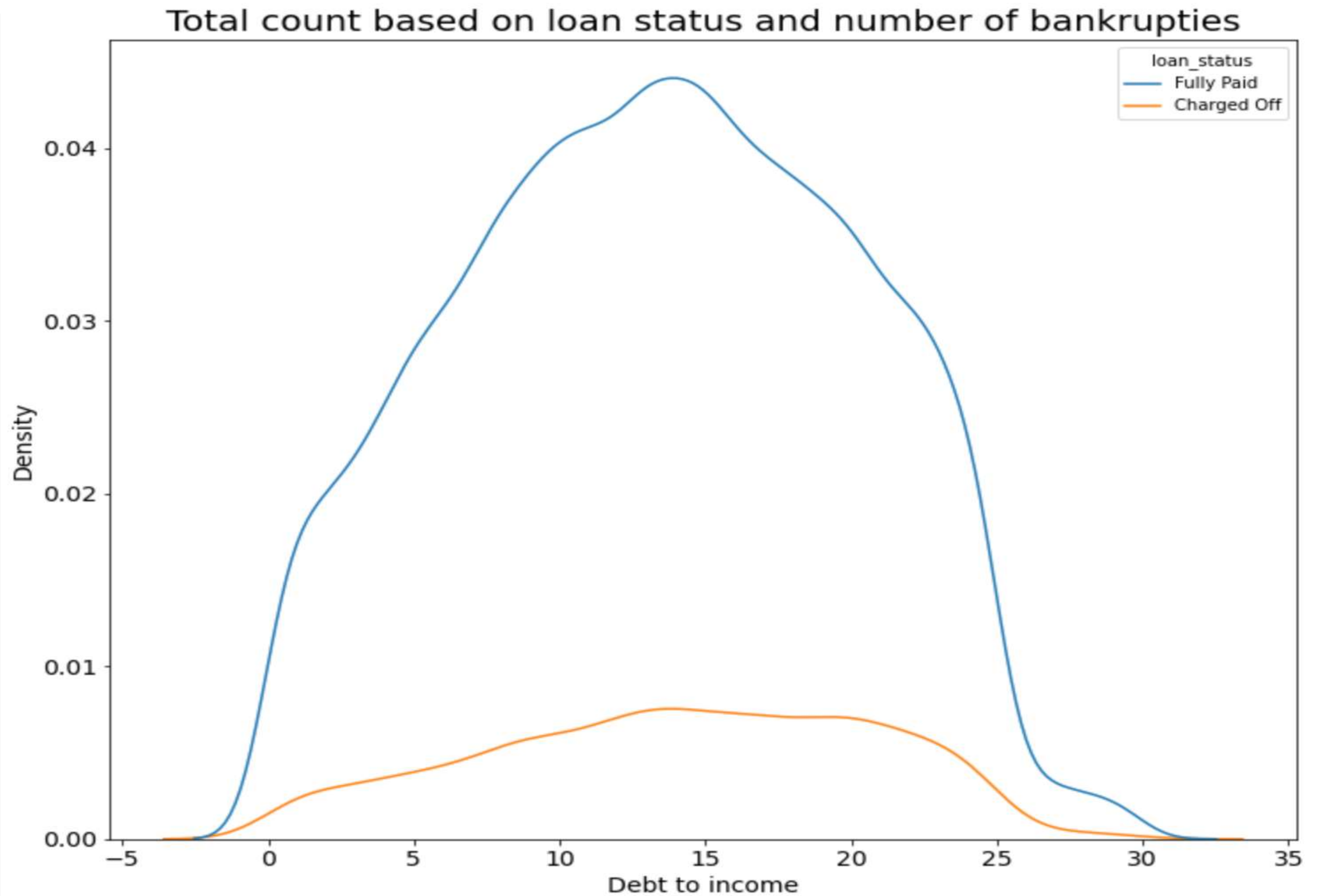


10.Debt to income:

Conclusion drawn about Debt-to-Income ratio

1.Majority of the borrowers have a dti ratio of 13.5

2.Borrowers whose dti ratio is between 10 and 15 is having more chance of defaulting.

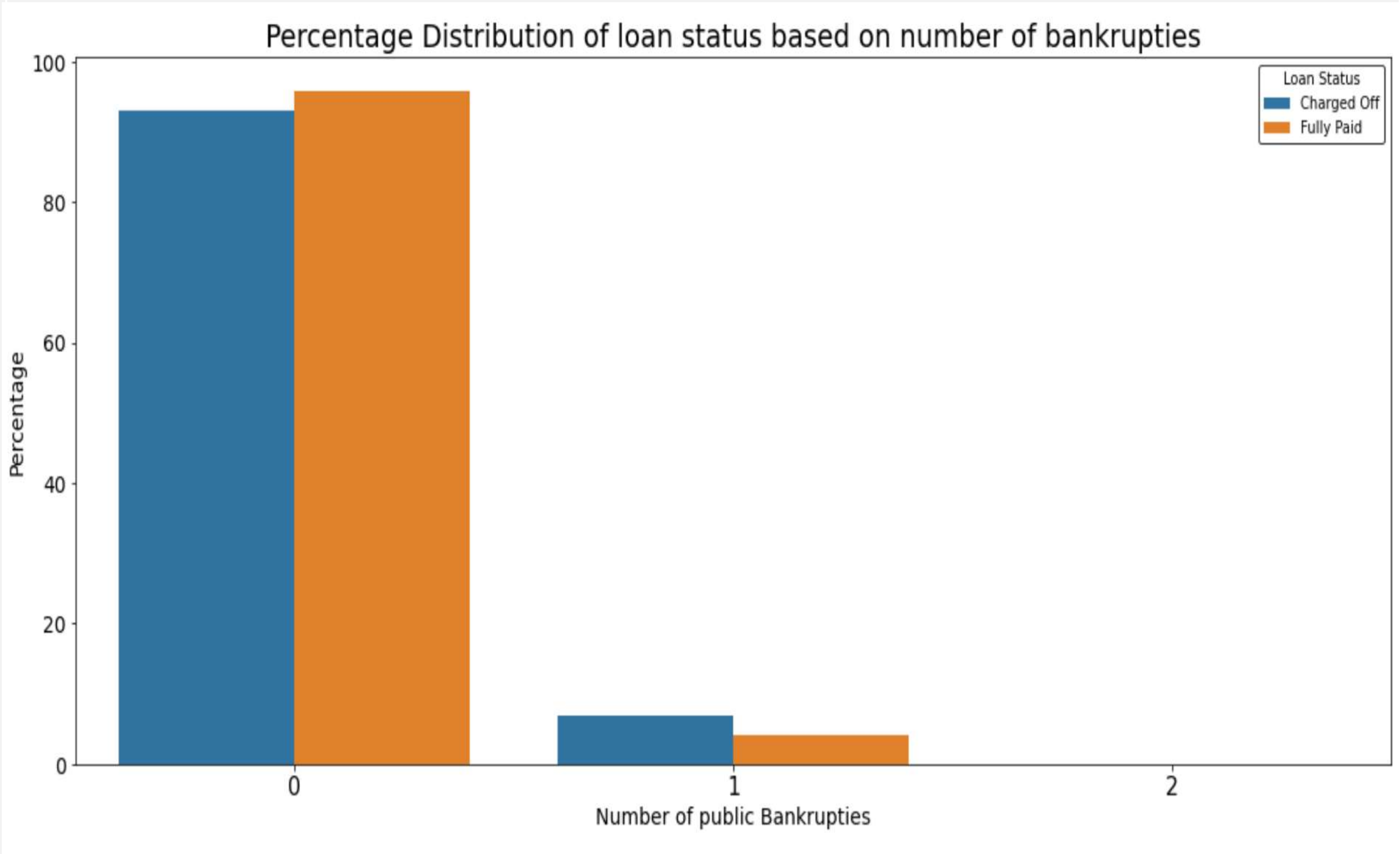


11.Public records bankruptcies:

Conclusion drawn about Number of bankruptcies:

1.Most of the borrowers have no recorded bankruptcies.

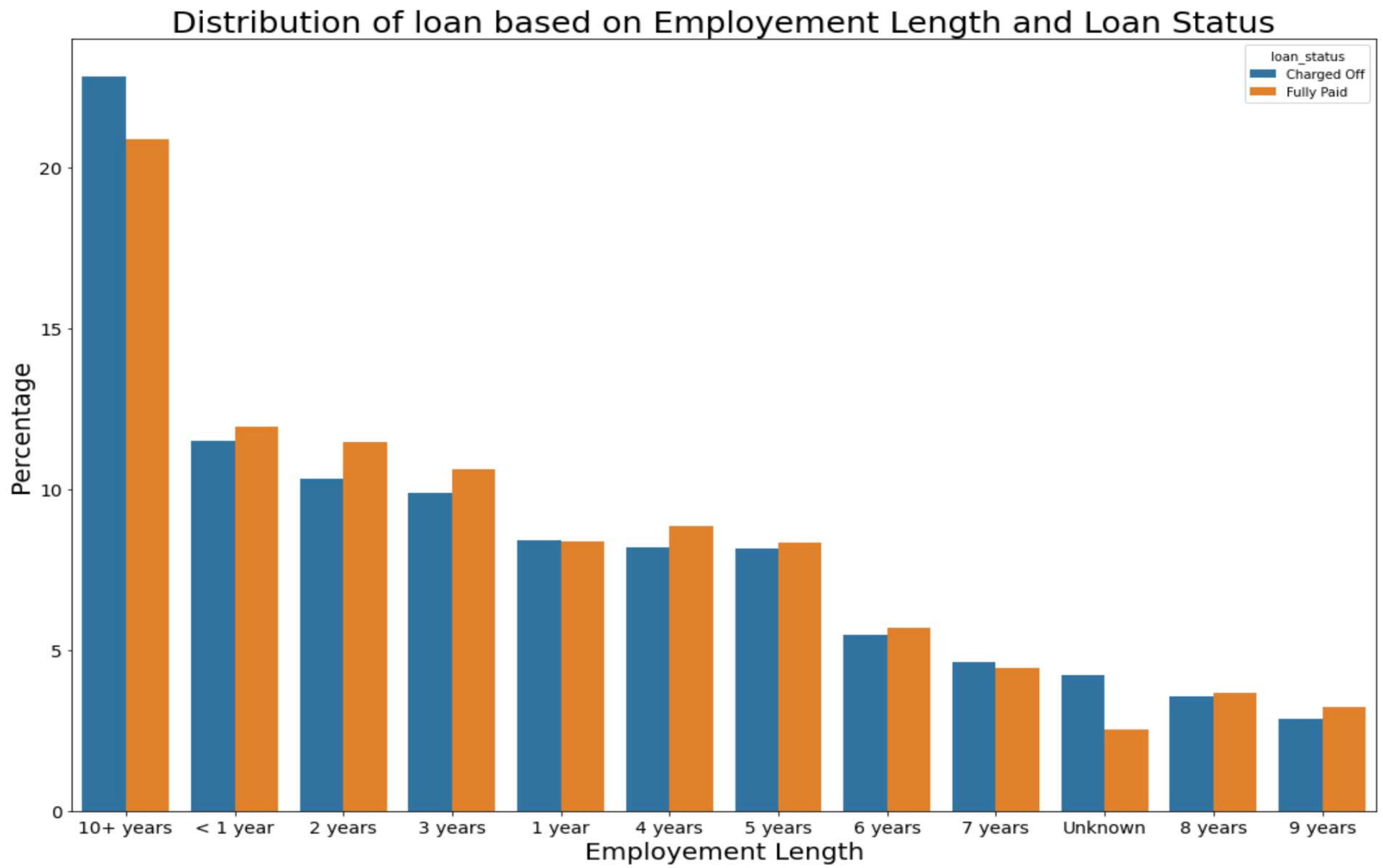
2.The chance of defaulting significantly increases if the borrower is having at least 1 bankruptcy.



13.Employment Length:

Conclusion drawn about Employment Length:

1.Based on years of employment length one cannot predict if the borrower is going to default.



FINAL CONCLUSION

Recommendations:

1. The company should avoid giving loan to borrowers whose income is below $40000 and who are requesting a loan amount greater than $20000.

2. The company should restrict borrowers who are belonging to Low Income Category and staying in Rent or Mortgage property and requesting loan greater than $15000 for the purpose of consolidation of debt.

3. Borrowers belong to Grade C,D,E have higher chance of defaulting. Company should be careful before sanctioning loan to such borrowers depending upon loan amount.

4. Apart from debt consolidation, borrowing taking high value loans for the purpose of credit card payment, small business and education and having less income have more chance of defaulting.

5. Borrowers belonging to state CA,FL, NY have high chance of defaulting. Hence borrowers belonging to these states having low income have high chance of defaulting based on Loan amounts. Such borrowers needs to be restricted.

6. The company can increase its profits in states having high number of High Income group borrowers by charging more interest rate. The states are NJ,VA,DC,MD.

7. For borrowers taking loan for renewable energy the company can further reduce interest rates to attract more customers.

Additional Insights

Below are some of the additional interesting insights obtained after analyzing the driver variables.

1.We could observe that many borrowers have been given loan at 7.5% interest rate. Beyond the 7.5% value there is sharp decline in number of borrowers given loan above 7.5%. One of the reason may be that beyond 7.5% interest rate additional taxes are levied.

2.Borrower belonging to High income category are taking high value loans for renewable energy. This might be due to high subsidy by government to push renewable energy utilization

3.Based on the data we can say that post recovery after the economic recession (year 2007-2009), large number of loans were taken in order to pay off the debts accumulated during the years of recession.

THANK YOU