LENDING CLUB CASE STUDY

Exploratory Data Analysis
Assignment

By

Koteswara Rao Annavarapu

koteswara.rao.aiml@gmail.com

APPROACH

- Business Understanding
- Problem Statement
- Data Understanding
- Data Cleaning
- Data Analysis Approach
- Key Recommendations

BUSINESS UNDERSTANDING

Lending Club is consumer finance company which specializes 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.
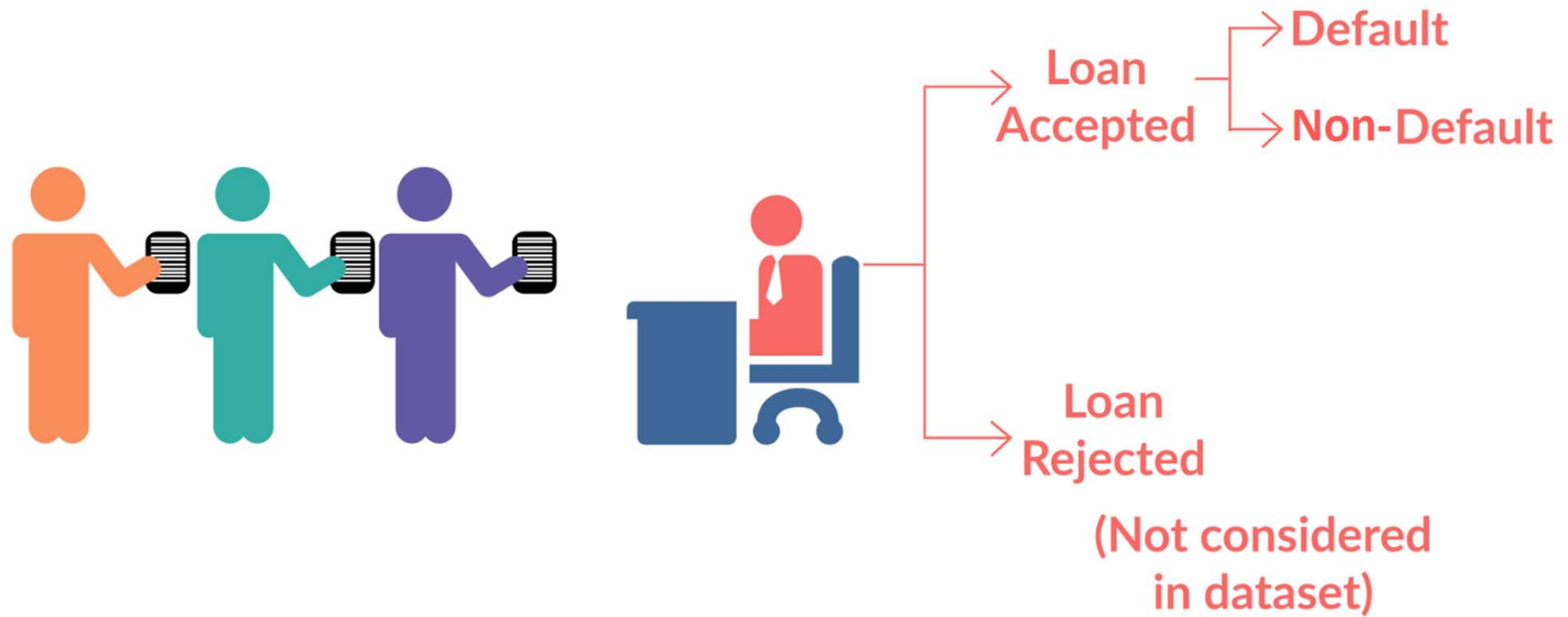
Two types of risks are associated with the bank's decision:

➢ 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 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.

LOAN DATASET

Loan Accepted → **Default**

Loan Accepted → **Non-Default**

Loan Rejected

(Not considered in dataset)

APPROACH

- Exploratory Data Analysis techniques are used to understand how **consumer attributes** and **loan attributes** influence the loan default tendency.
- For this I have performed the data cleaning as first step
  - Dropped all the columns with >50% of null values
  - After dropping the null value columns , studied the metadata of remaining 57 columns with definitions from Data dictionary and
  - Later performed fixing the column data to correct data type (from object type to respective float or string or date types)
  - Cleaning of the data to suit the correct formats like cleaning extra characters in "10,000" , ">10 years", "10.5%" to convert it from string to a quantitative variables for analysis
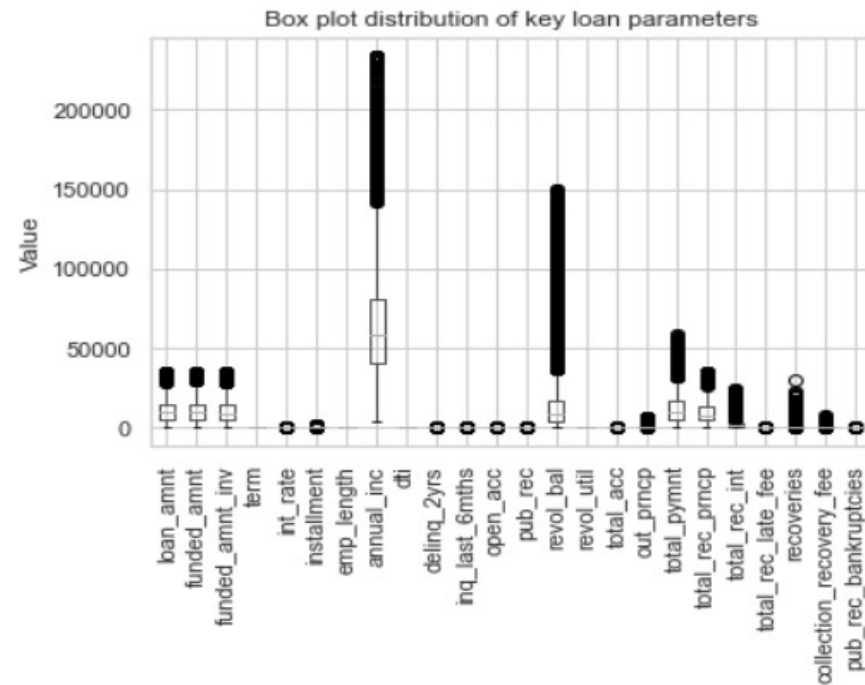
OUTLIER REMOVAL

➢ Performed outlier analysis and fixed the annual_income column which has clear indication of outliers. After dropping the rows at 99th percentile the distribution looks fine.

Data Distribution Before Cleanup
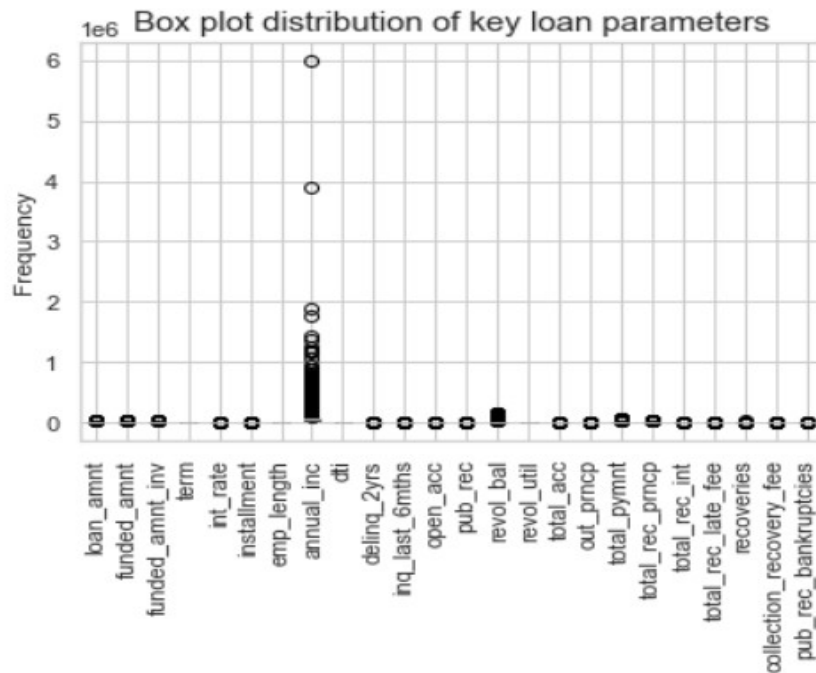
Data Distribution after outlier treatment

DATA ANALYSIS

- From the remaining dataset identified the key loan attributes and consumer attributes
- Differentiated them into categorical variables and Quantitative variables
- Performed Univariate analysis on all the possible variables
- For the purpose of determining the loan default behaviour, established "loan_status" as the target column
- Perfomed Segmented Univariate analysis on the loan_status dimensions like "Fully Paid" and "Charged Off(defaulted)" status
- Derived additional inferences by splitting the issued_d column into month and year .
- Finally, performed Bivariate Analysis on the "Loan Defaulters subset of dataset".

UNIVARIATE ANALYSIS – KEY OBSERVATIONS

- ➢ A total of 16% of loans issued by Lending club got defaulted(charged_off)
- ➢ Average size of loans issued were small loans between $5000 to $10000
- ➢ Average installment size is $200
- ➢ The average income of the borrowers between $50k to $60k
- ➢ Loan approval rate is high to borrowers with (0%) revolving credit utilization
- ➢ Increase of Revolving Credit balance decreases the chance of loan approval
- ➢ Majority of loans were given to employers with Debt to Income ratio of 10 to 20
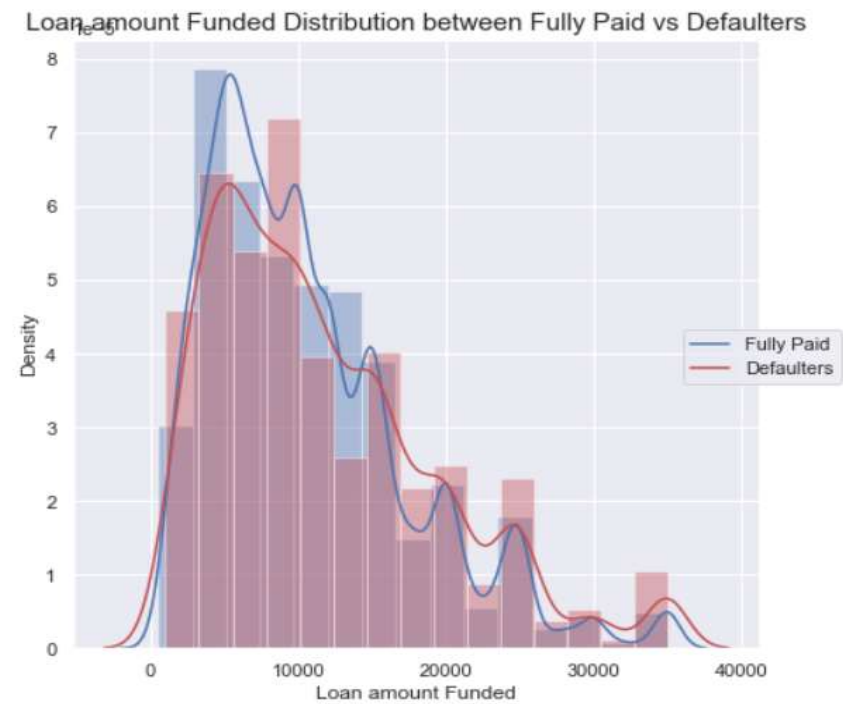
KEY RECOMMENDATIONS

➤ Lending Club issuing majority of loans without Veritication should as it might lead to loan defaults.
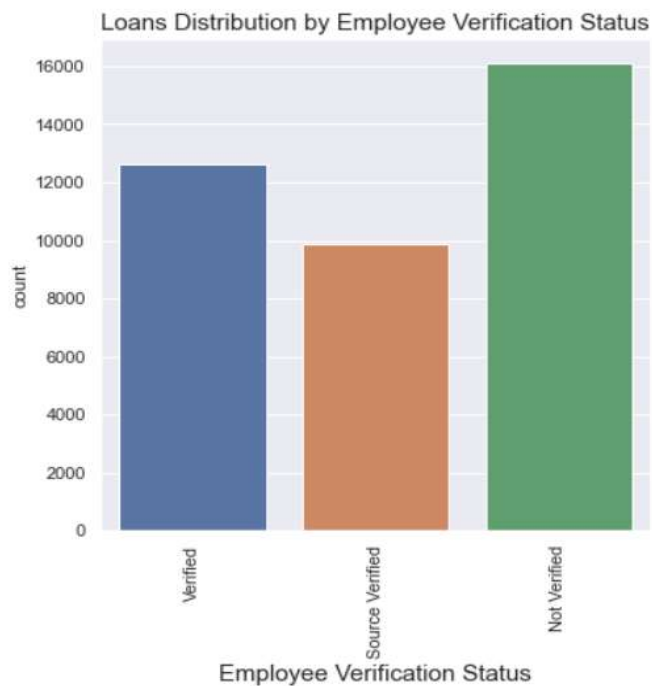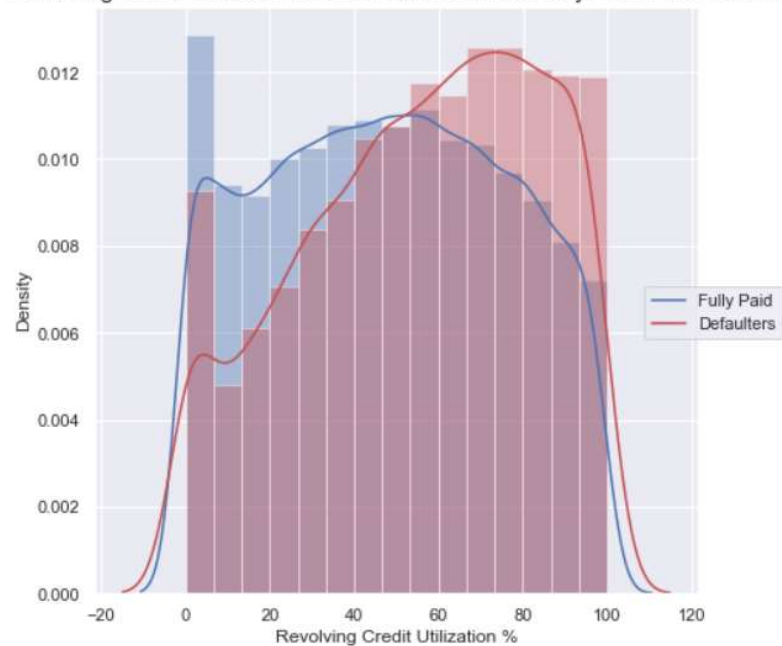
➤ Higher the Loan amount the more is chance of default and recoveries.

Loans Distribution by Employee Verification Status

Loan amount Funded Distribution between Fully Paid vs Defaulters

KEY RECOMMENDATIONS

➤ Lending club should avoid lending to people with above 50% revolv_credit_utilization

➤ Lending at higher interest rates increases the chance of loan default.

Revolving Credit Utilization % Distribution between Fully Paid vs Defaulters

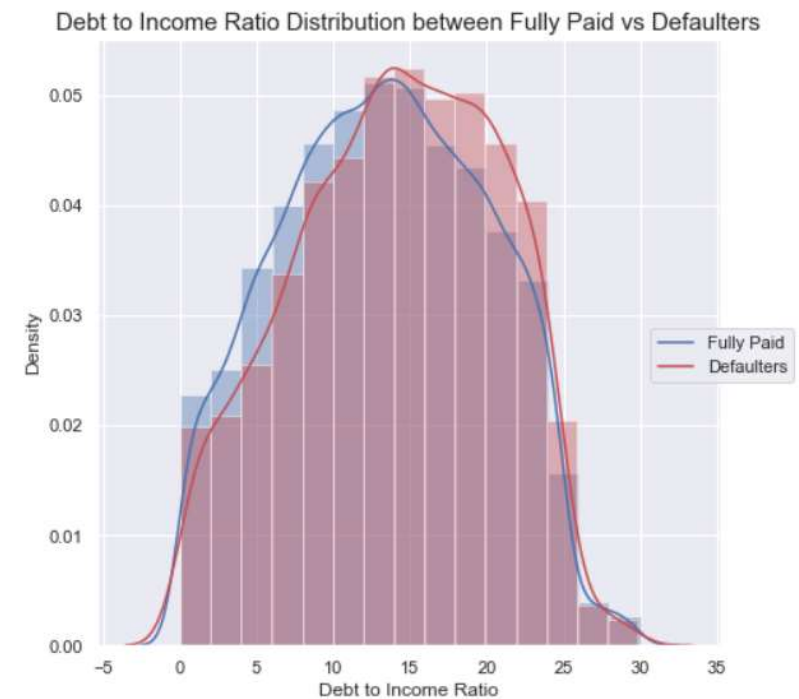Interest Rate Distribution between Fully Paid vs Defaulters
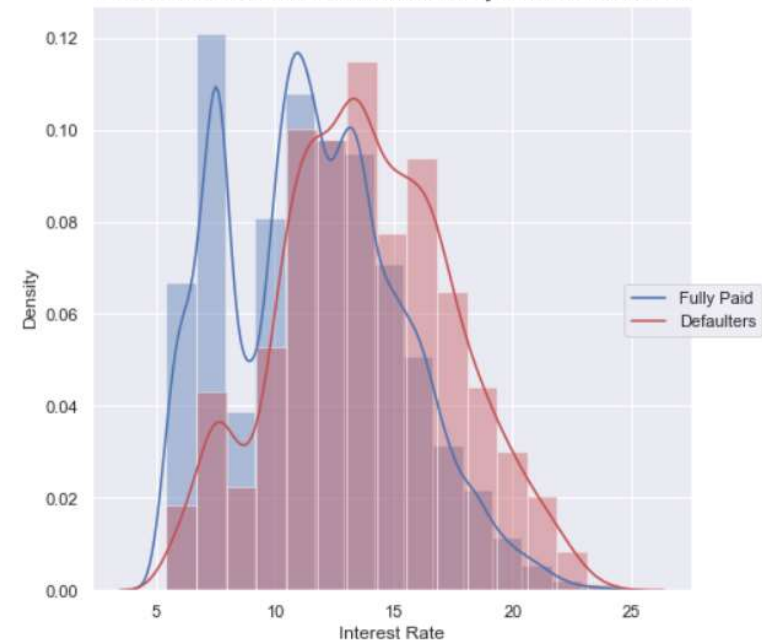
KEY RECOMMENDATIONS

➢ Lending at interest rate above 14% has higher chances of default
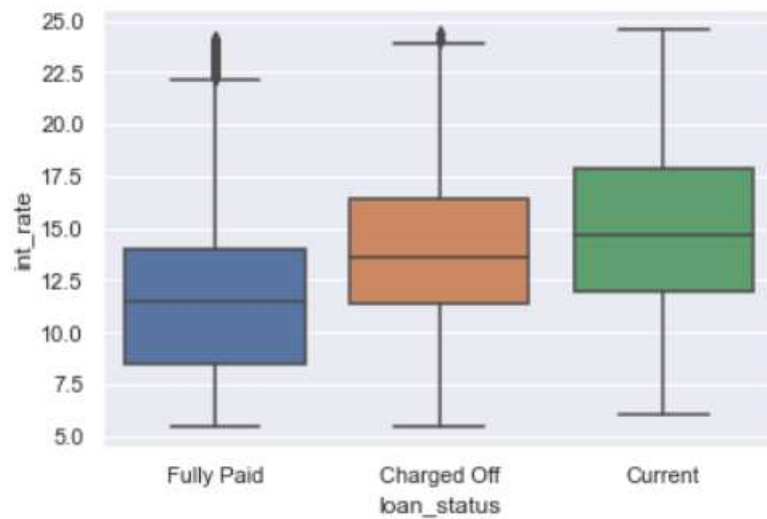
➢ Debt to Income ration of the borrower should preferably be <= 15 to avoid defaults.

Debt to Income Ratio Distribution between Fully Paid vs Defaulters

KEY RECOMMENDATIONS

➢ Loans should preferably be not given to people with Public Insolvency Records

Public Insolvency Records Distribution between Fully Paid vs Defaulters
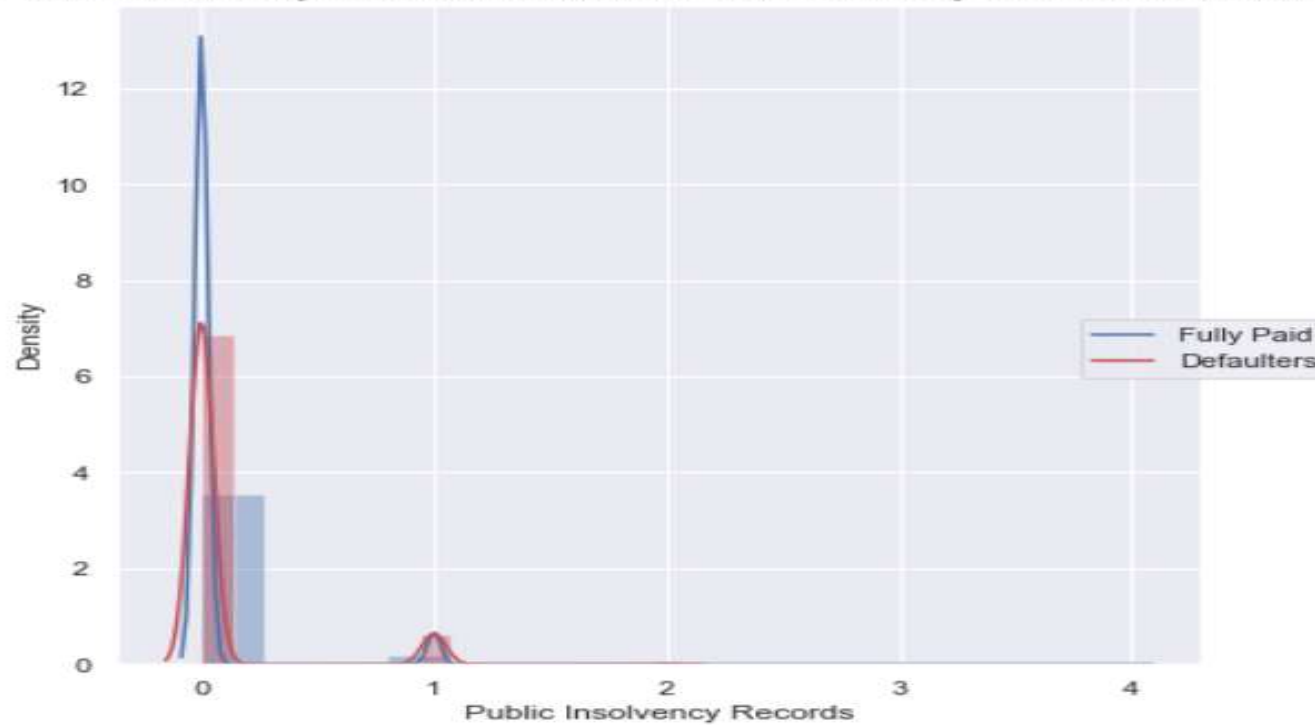
CAUTIONS TO BE EXERCISED

➢ LC Loan grades B , C , D should undergo extra scrutiny for following subgrades in the order B5 , B3, B4 , C1 , C2

Loan Grade Distribution between Fully Paid vs Defaulters

CAUTIONS TO BE EXERCISED

➢ LC Loan grades B , C , D should undergo extra scrutiny for following subgrades in the order B5 , B3, B4 , C1 , C2

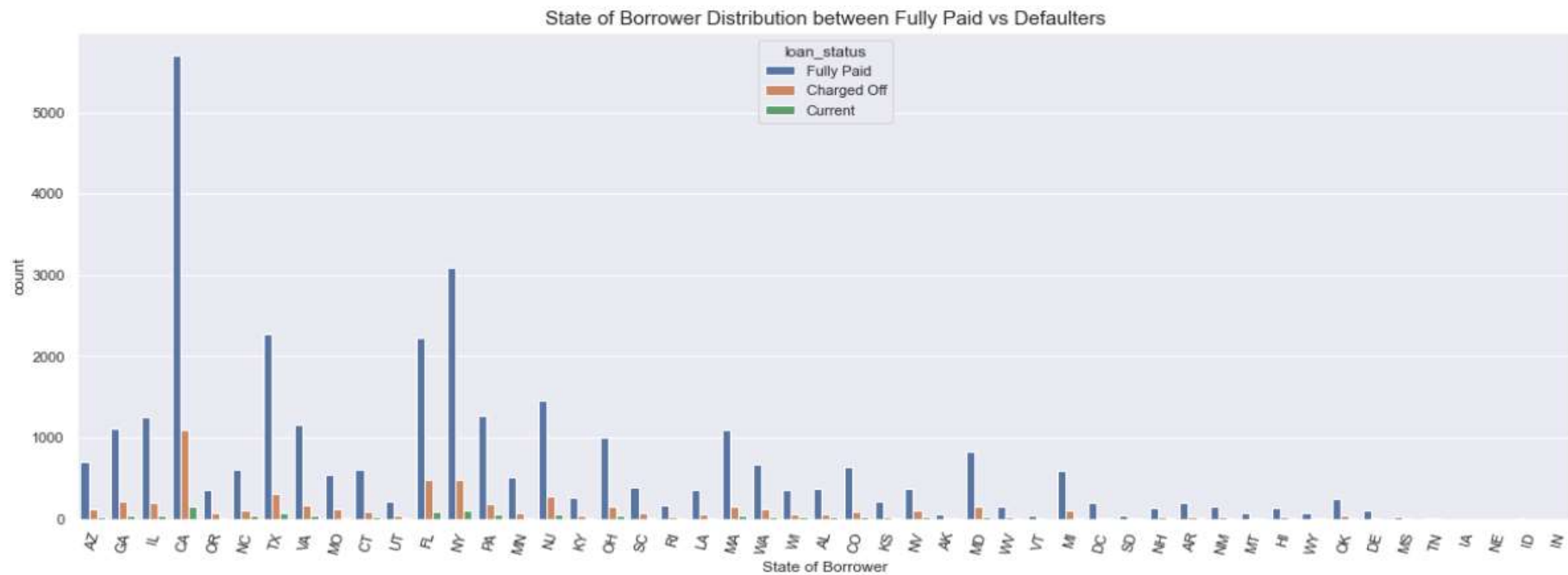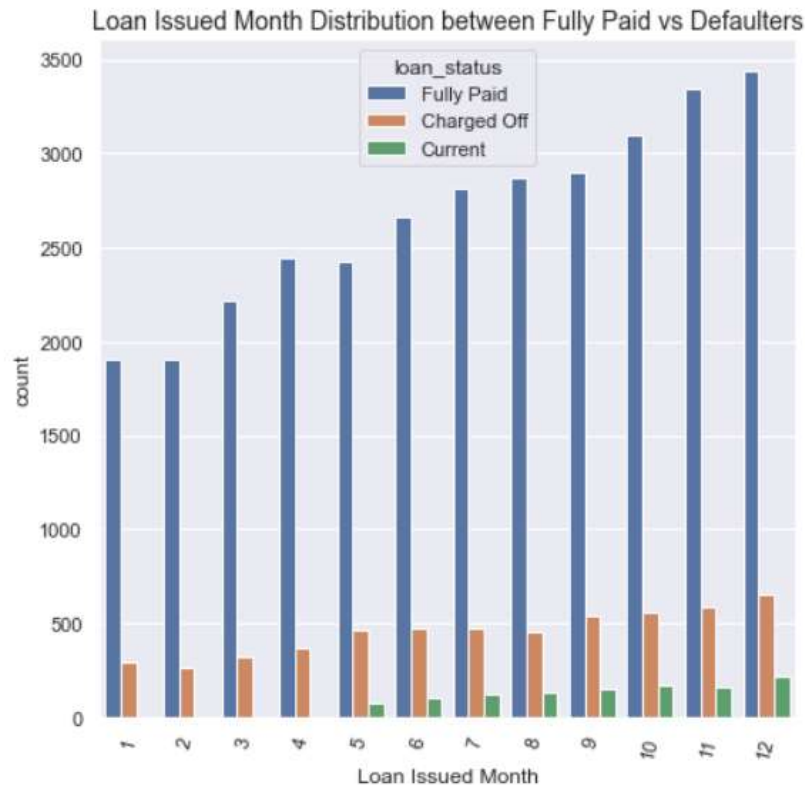Loan Sub Grade Distribution between Fully Paid vs Defaulters
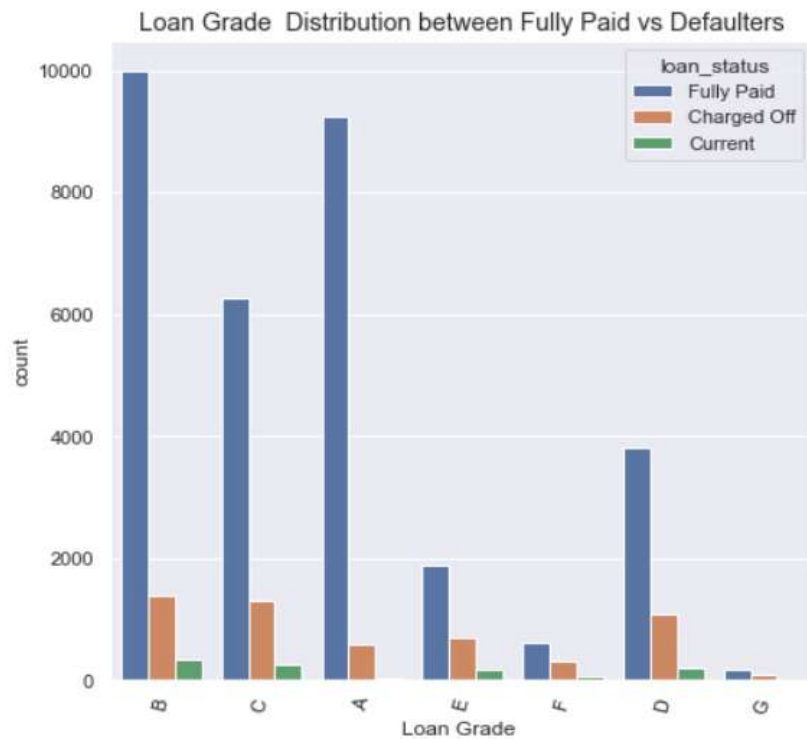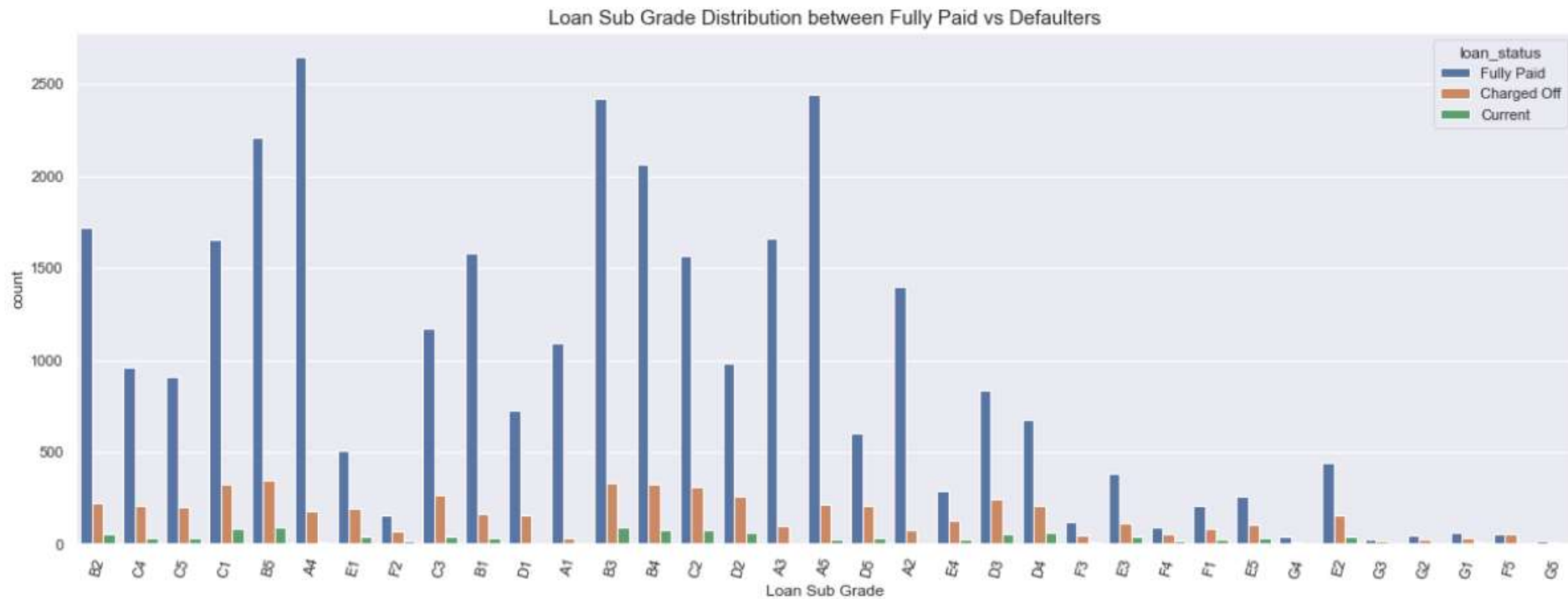
CAUTIONS TO BE EXERCISED

➢ Loan defaults are higher in these four states
1.CA California 2. FL Florida
3. NY Newyork 4. TX Texas

State of Borrower Distribution between Fully Paid vs Defaulters

CAUTIONS TO BE EXERCISED

Loan Issued Month Distribution between Fully Paid vs Defaulters

➢ Loans issued in last quarter of the year have higher chance of defaults compared to other months. One inference could be to meet the year end targets loans are issued at faster pace with easier approvals

RECOMMENDATIONS FOR LOAN APPROVALS

➢ Chance of loan default keeps decreasing with increase of Employee Annual Income

Employee Annual Income Distribution between Fully Paid vs Defaulters