

LENDING CLUB CASE STUDY SUBMISSION

By: Dhruv Sharma

Problem Statement

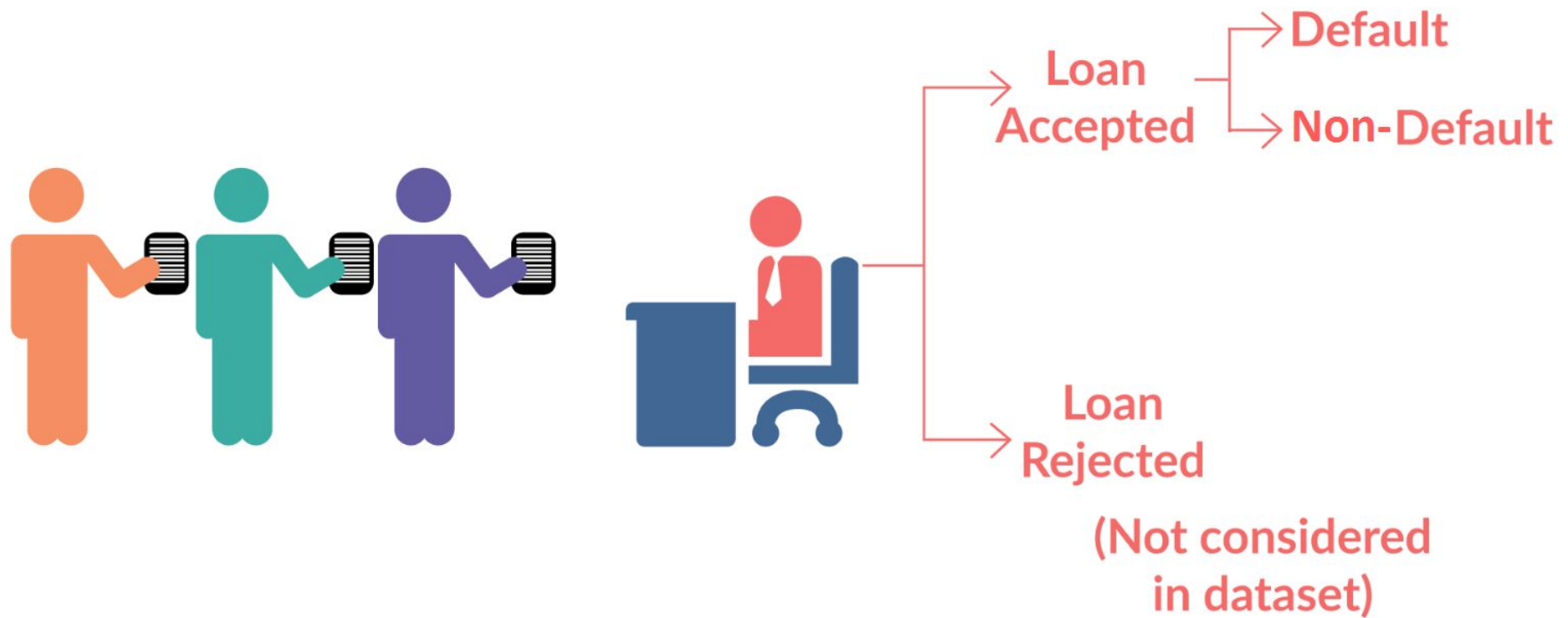
The problem statement assumes we work for a consumer finance company which specialises in lending various types of loans to urban customers. When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile. Two types of risks are associated 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 dataset given 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, we will use EDA to understand how consumer attributes and loan attributes influence the tendency of default.

LOAN DATASET



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)

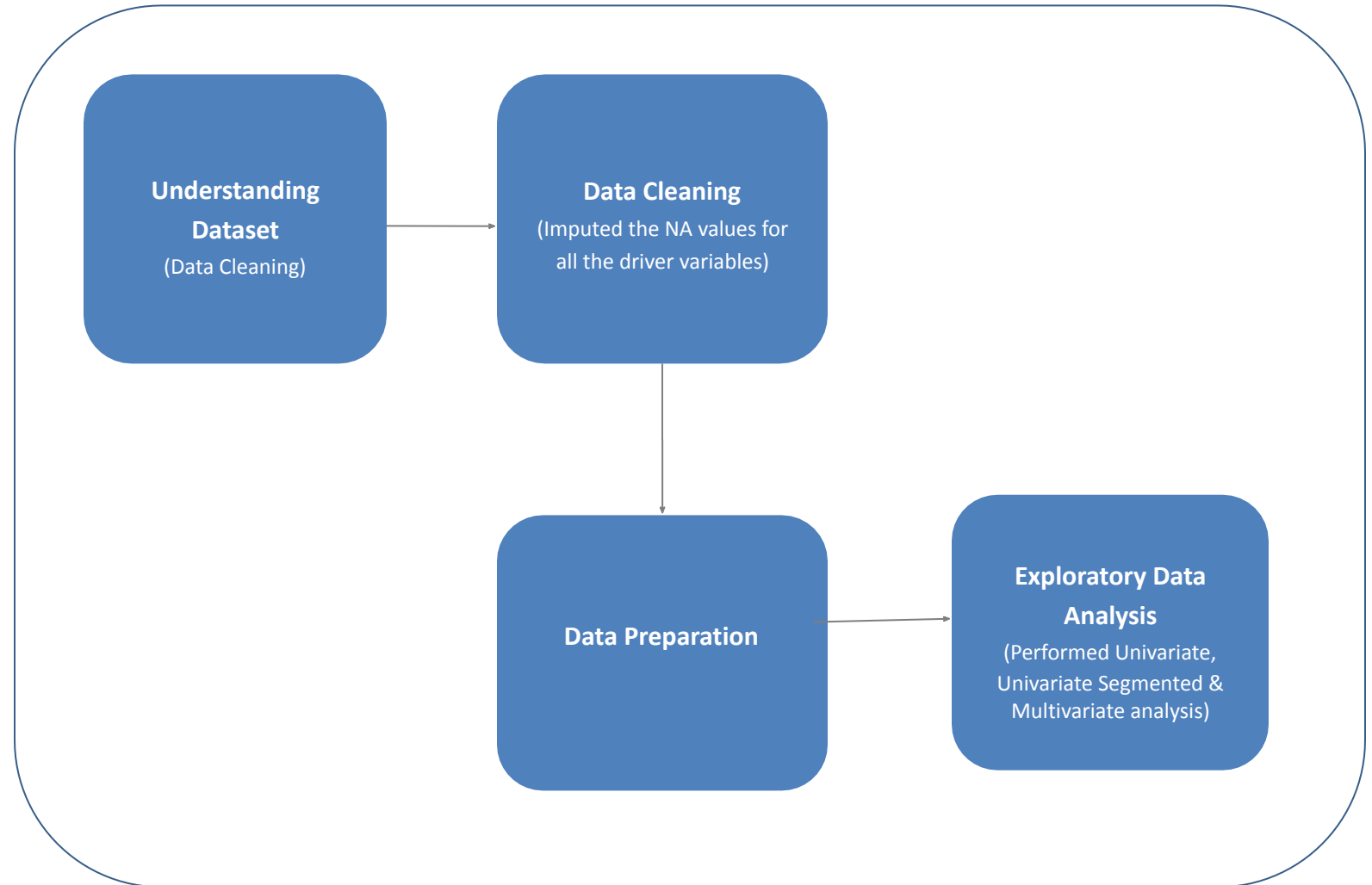
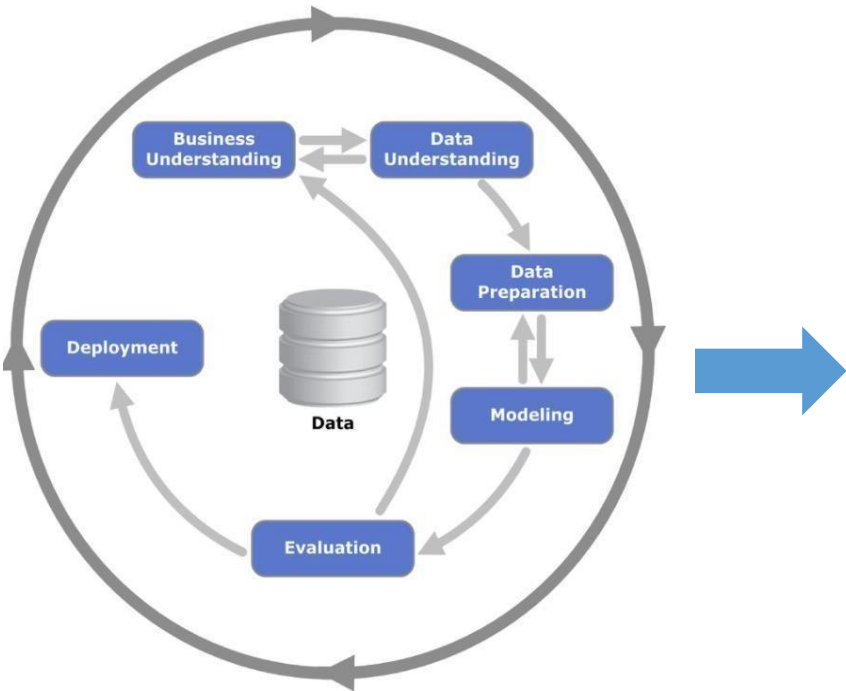
Case Study Objectives

Identification of Loan Applicant traits that tend to be risky applicants as they are more likely 'defaulters'

To differentiate b/w applicants that can pay loans without defaulting and other risky applicants to optimise the lending business

Loan Lending Organizations may choose to utilize this knowledge for its portfolio and risk assessment of new loan applicants

Problem solving methodology using CRISP-DM



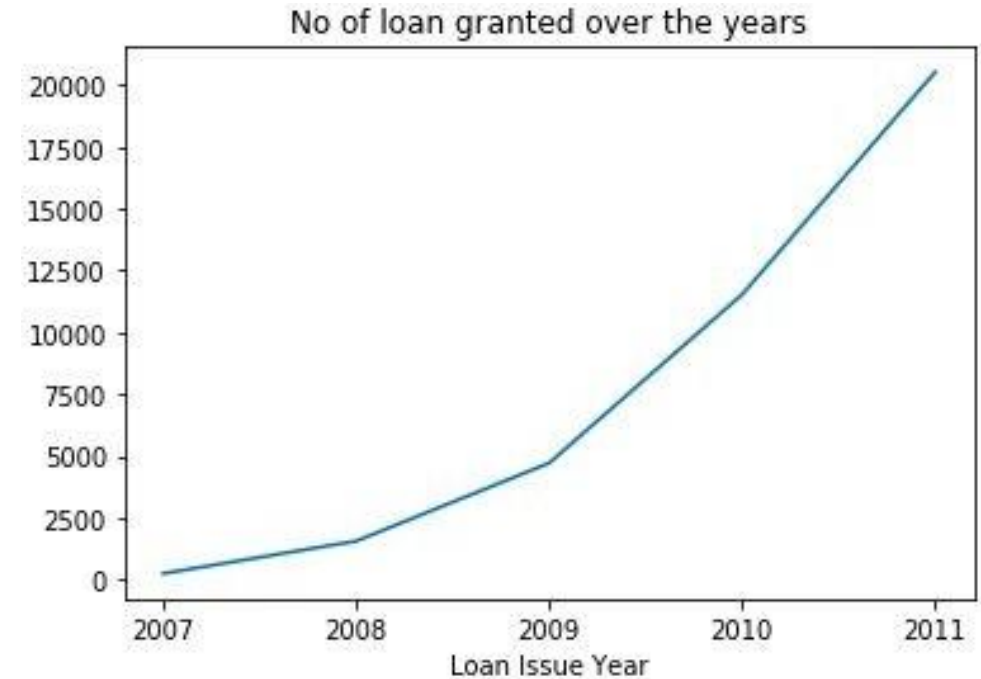
Data Cleaning Steps

- **Delete columns:** Delete unnecessary columns.
- **Remove outliers:** Remove high and low values that would disproportionately affect the results of your analysis.
- **Missing values:** Treat missing values with appropriate approach.
- **Duplicate data:** Remove identical rows, remove rows where some columns are identical.
- **Filter rows:** Filter by segment, filter by date period to get only the rows relevant to the analysis.

Analysis

- The essence of the whole project is to analyze and understand how consumer attributes and loan attributes are influencing the tendency of defaulting.
- We performed **data cleaning and preparation** on the Loan dataset:
 - Imputed the NA values for all the variables
 - Created two new columns:
 - Profit and Loss column
 - Ratio of funded amount and annual income
- During **univariate analysis** we have created:
 - Histograms and Bar charts to check out the distribution of all the driver variables
 - Box plots to detect the Outliers
 - Performed the Multivariate analysis to understand how different variables interact with each other.

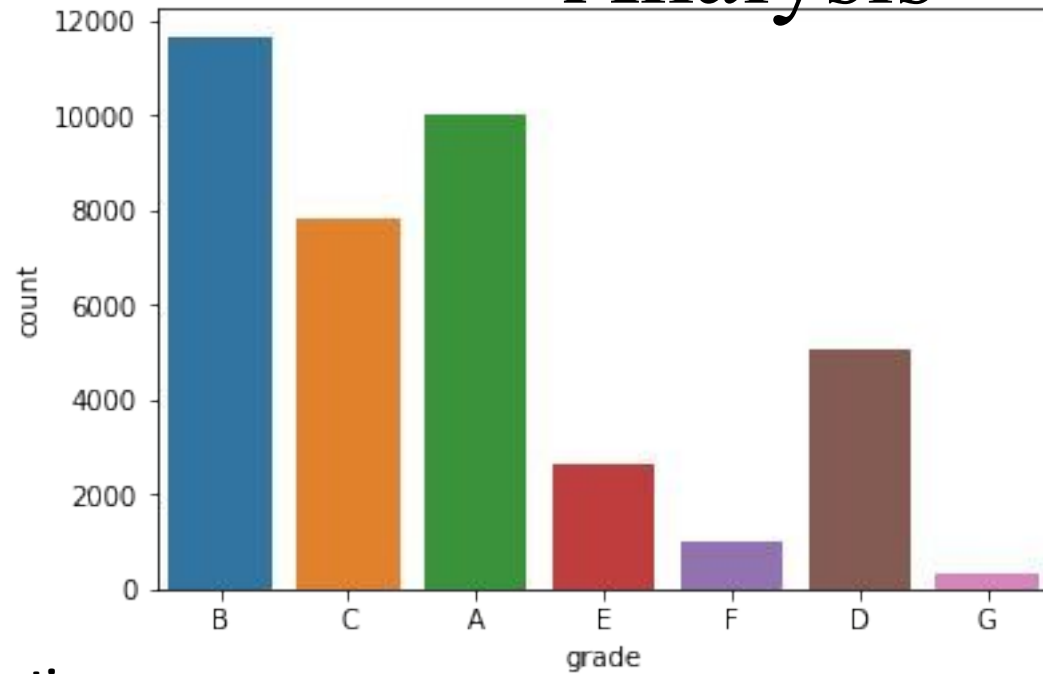
Loan Status Analysis



Observations:

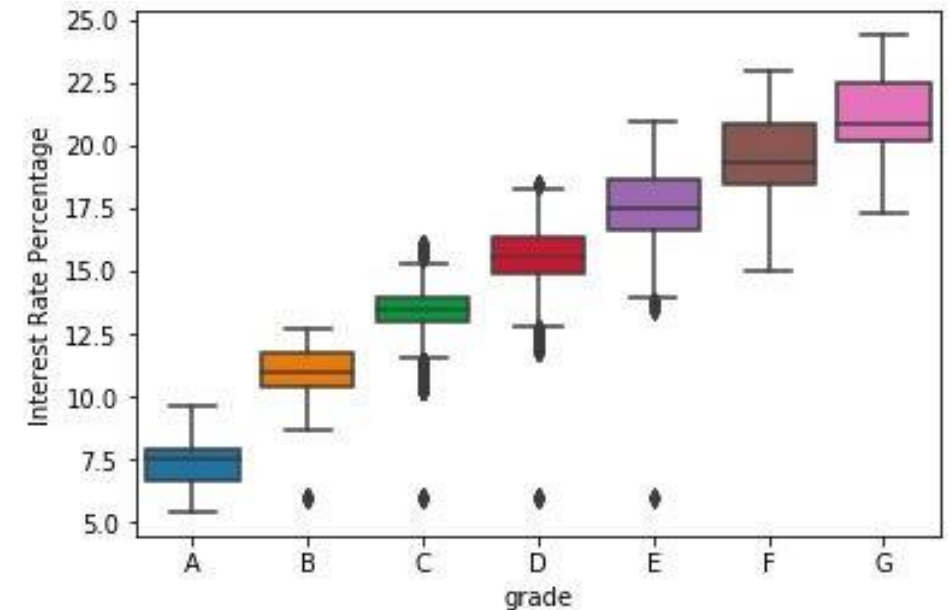
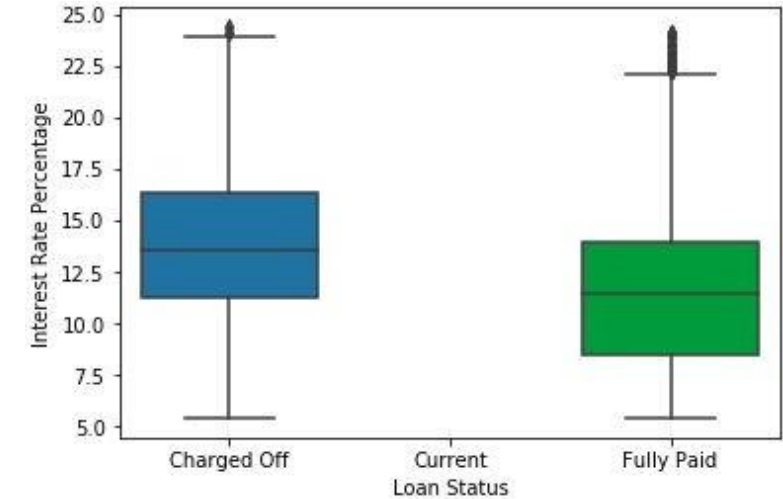
1. Most of the loans are Fully Paid.
2. About 14% of loan are having status as defaulters.
3. The loan has been increasing exponentially over the years.

Loan Grade Analysis

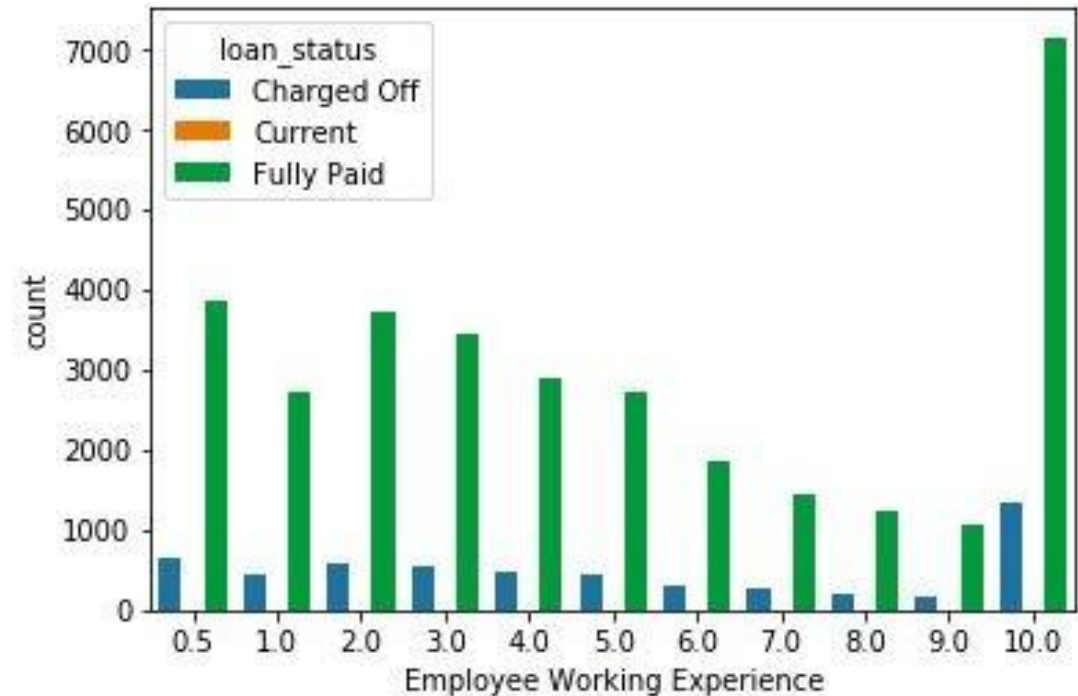
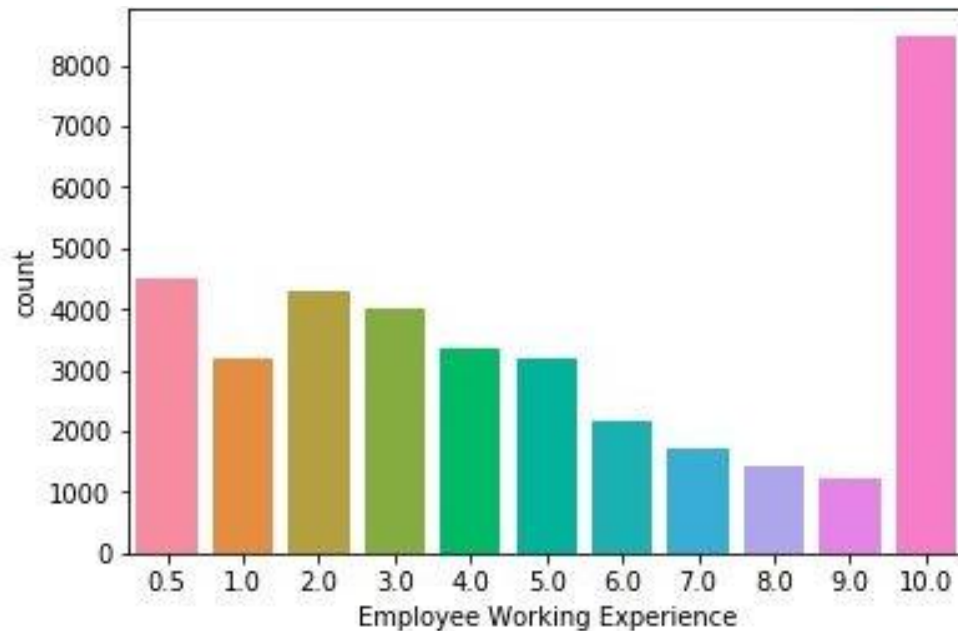


Observation:

1. Most of the loans have grade of A and B. Therefore stating most of the loans are high graded loans.
2. High Quality loans have low interest rate.
3. Which depicts, higher interest rate have higher tendency to default the loan.



Loan Applicants work Experience Analysis

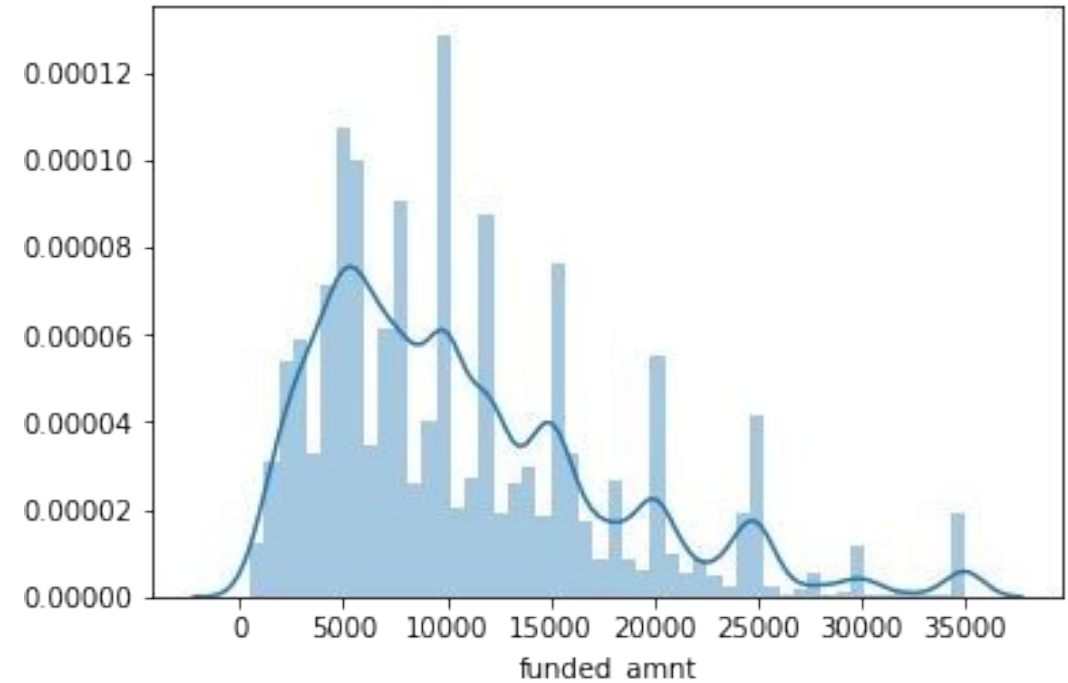
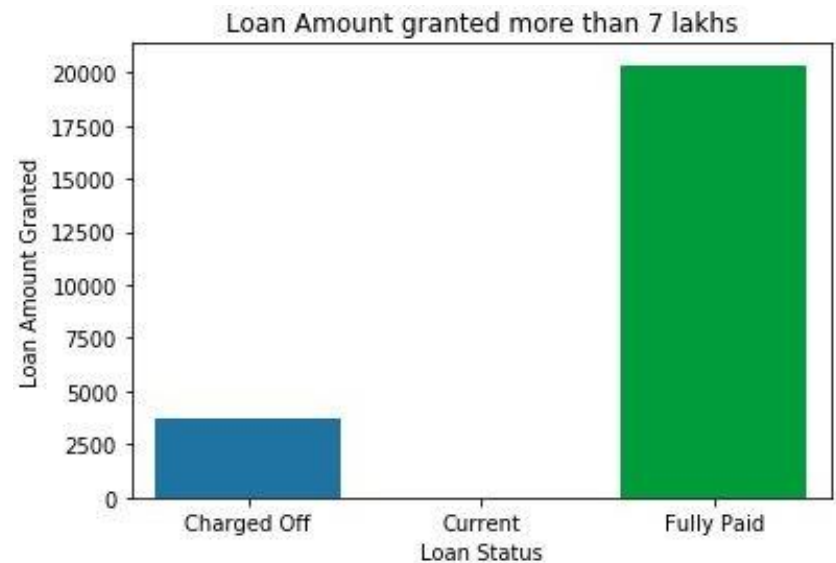
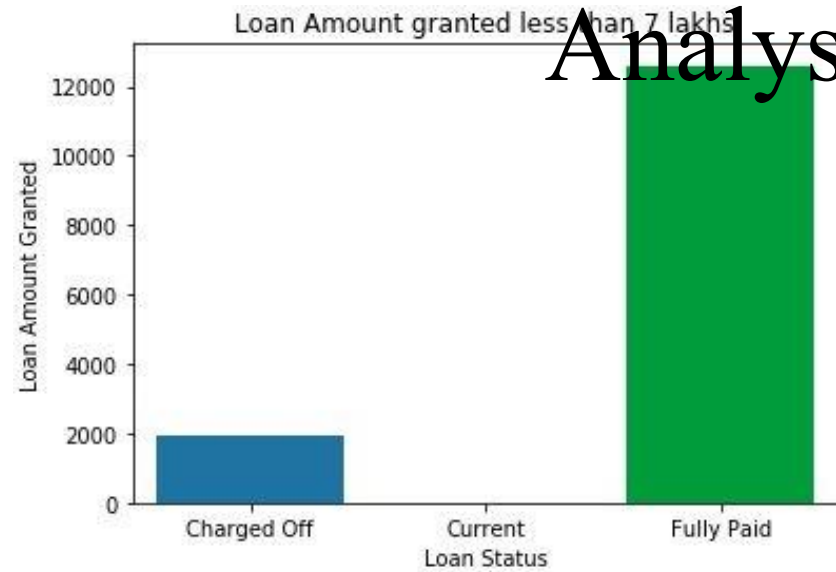


Observation:

1. Majority of employees applying for the loan have more than 10 years of experience
2. Tendency of person to default the loan with 10 years of experience is also high. So company need to be careful when granting loan.

Loan Amount

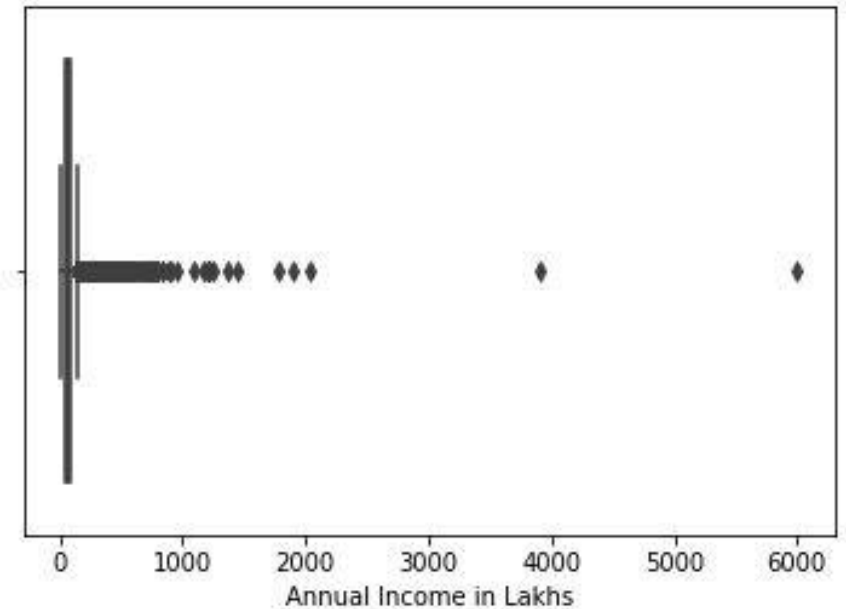
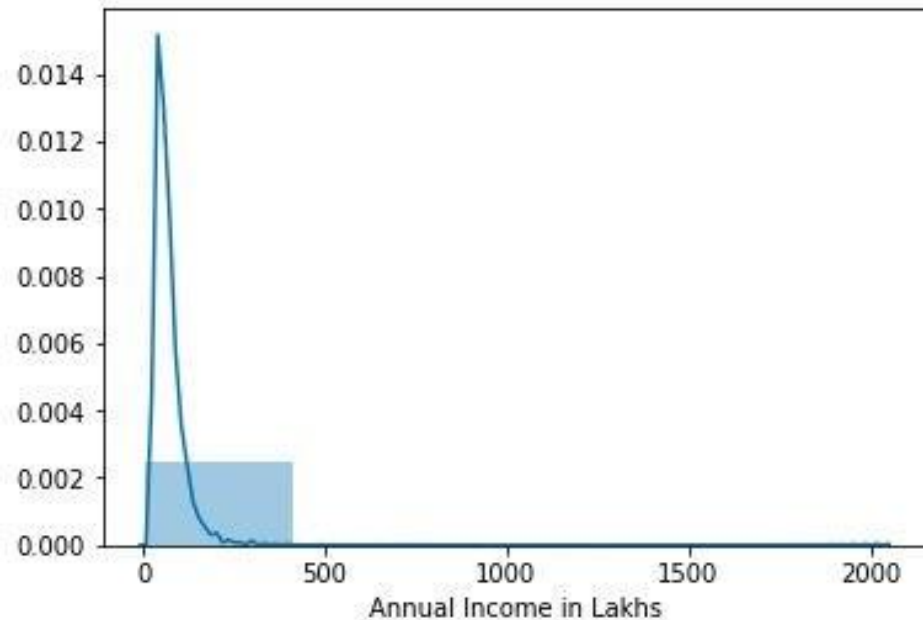
Analysis



Observation:

1. Funded amount is left skewed. Most of the loan amount given is below 7 lakhs.
2. Probability of people with loan amount greater than 7 lakhs tends to default is more than the people with less than 7 lakh

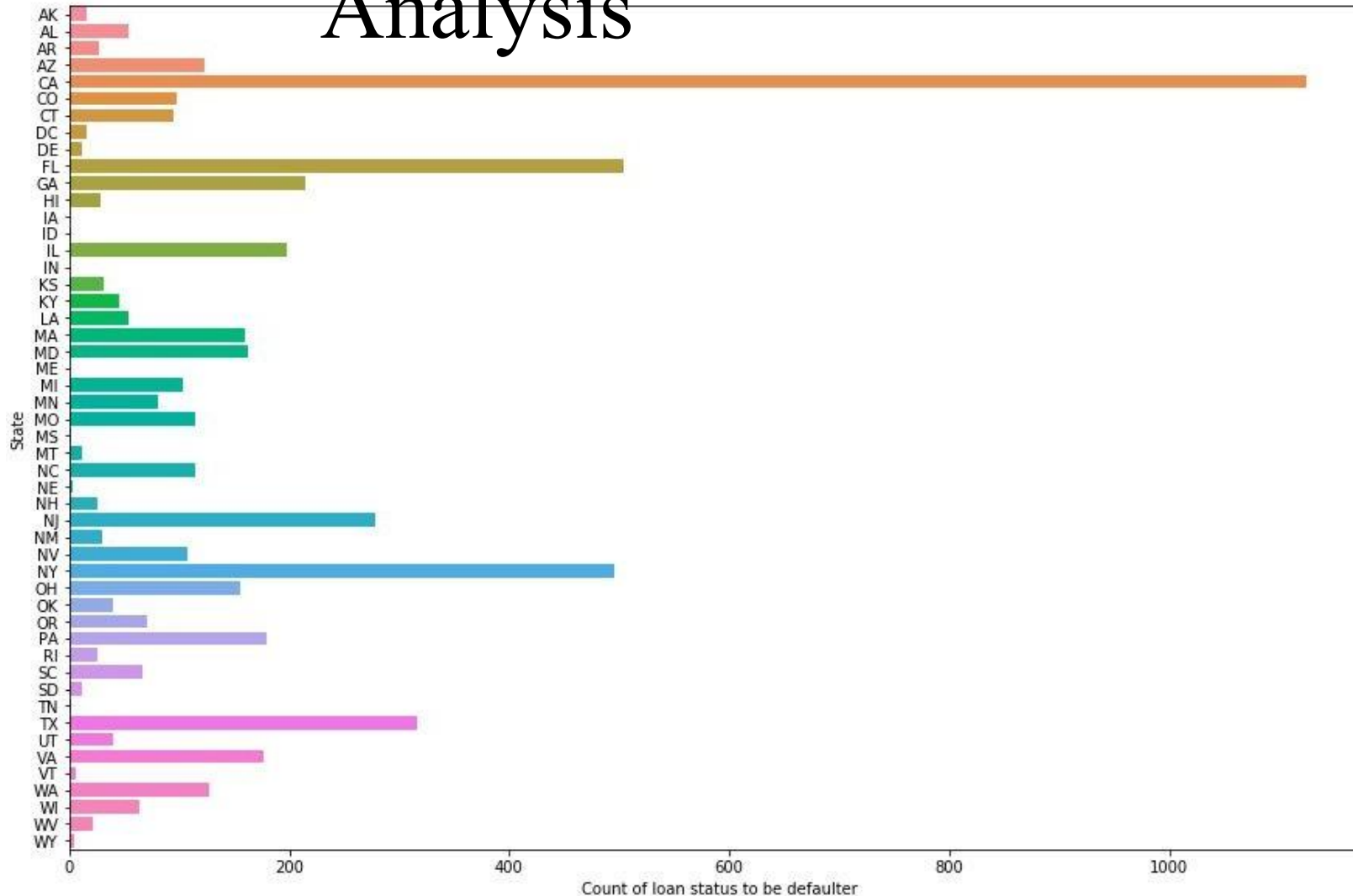
Income Analysis for the loan applicants



Observation:

1. Most of the applicants have income between 0 to 5 lakhs.
2. There are two applicants having income above 30 lakhs

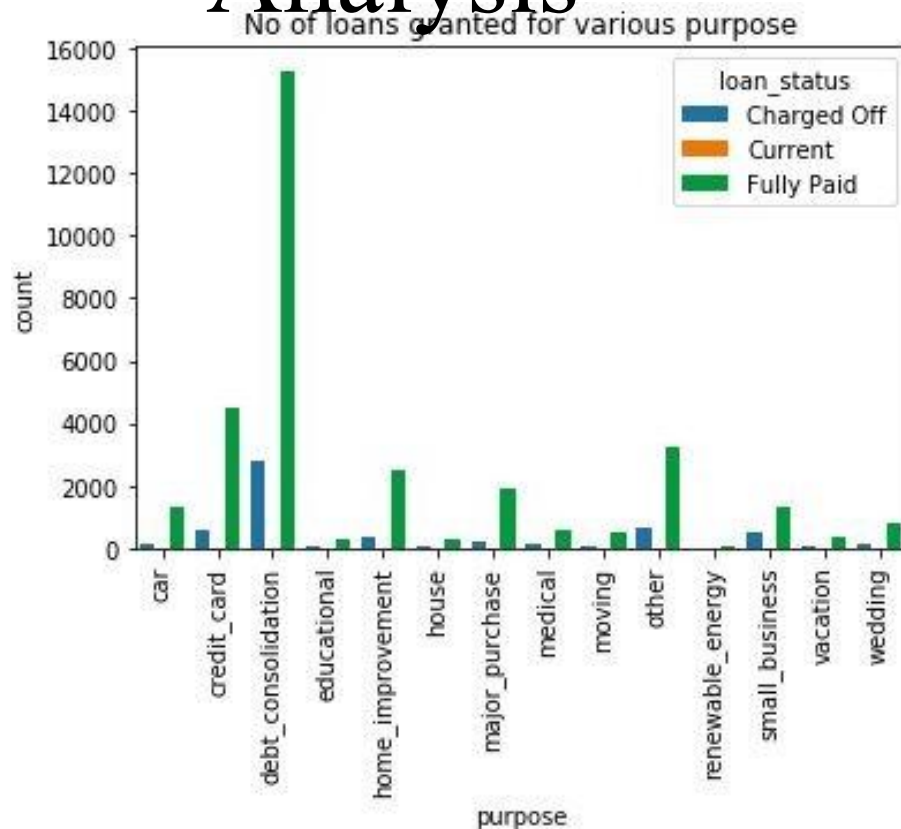
Loan Defaulters by State Analysis



Observation:

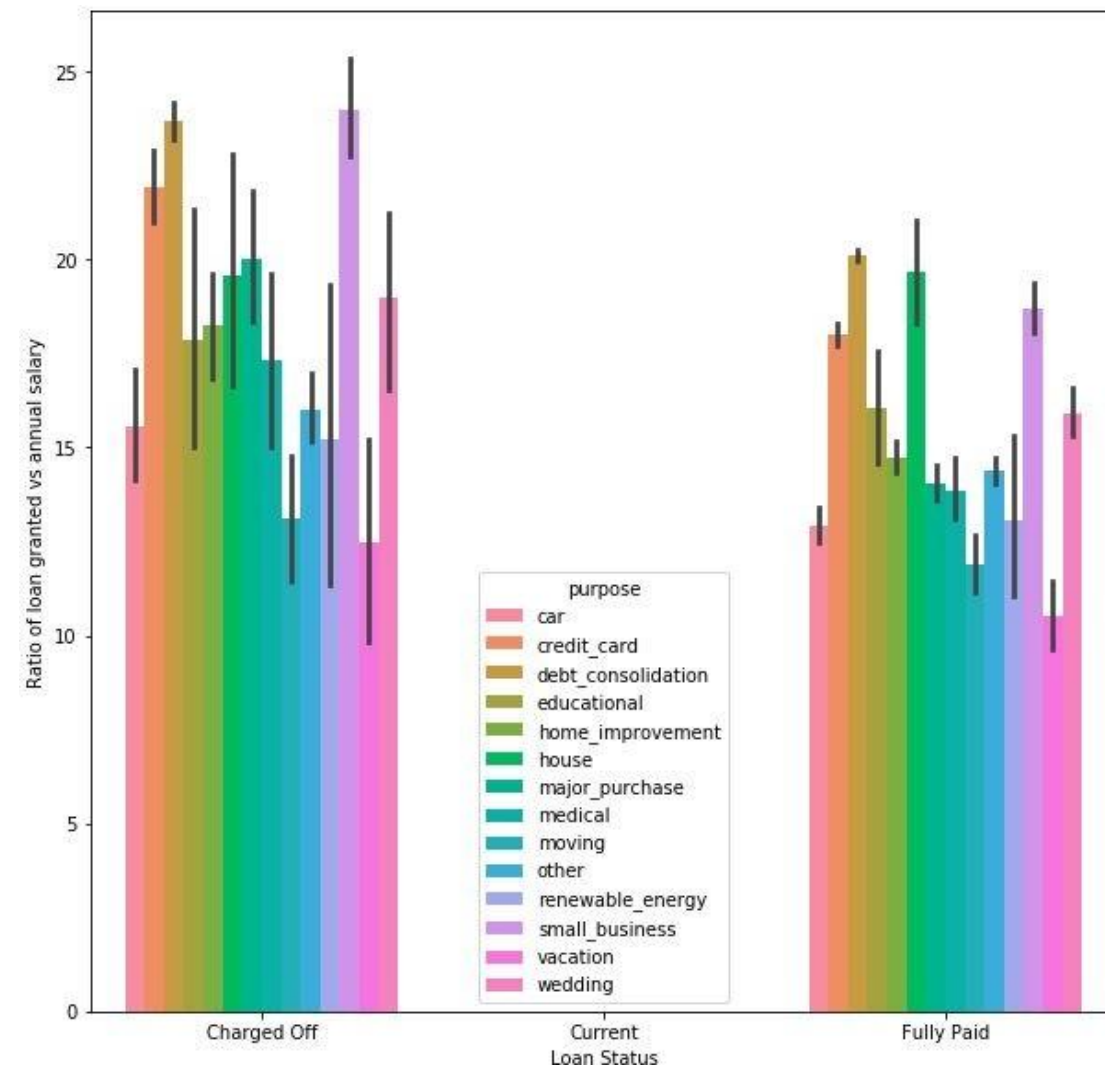
Applicants from state CA have high tendency to default the loan

Purpose for applying Loan Analysis

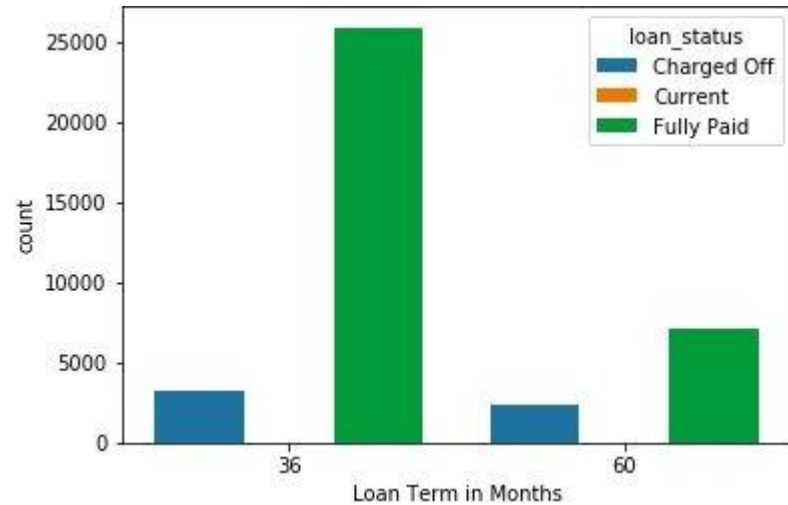


Observation:

1. Most of the loans are granted were for debt consolidation.
2. Applicants having higher income tends to default the loan.

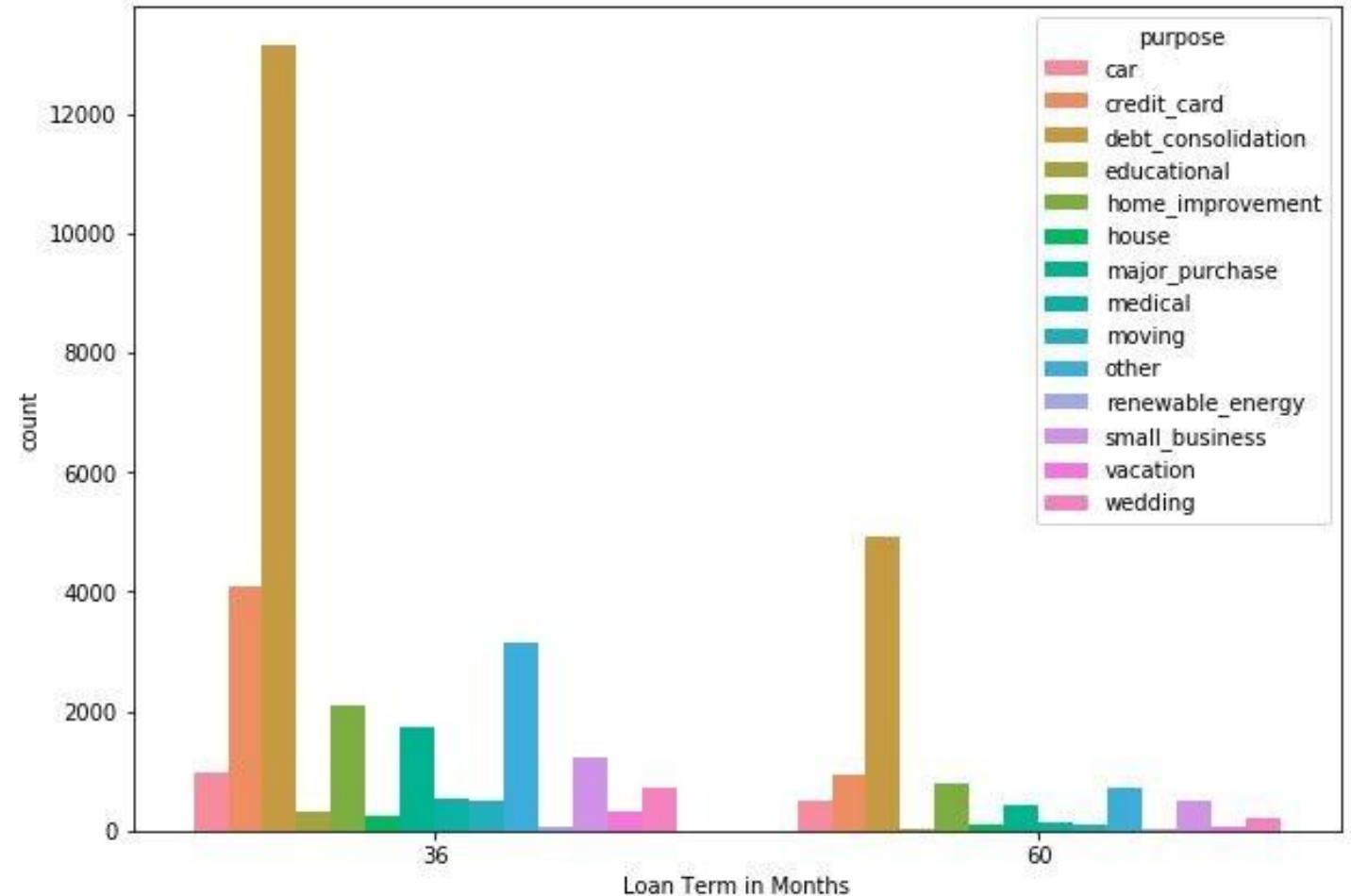


Loan Status vs Loan Tenure

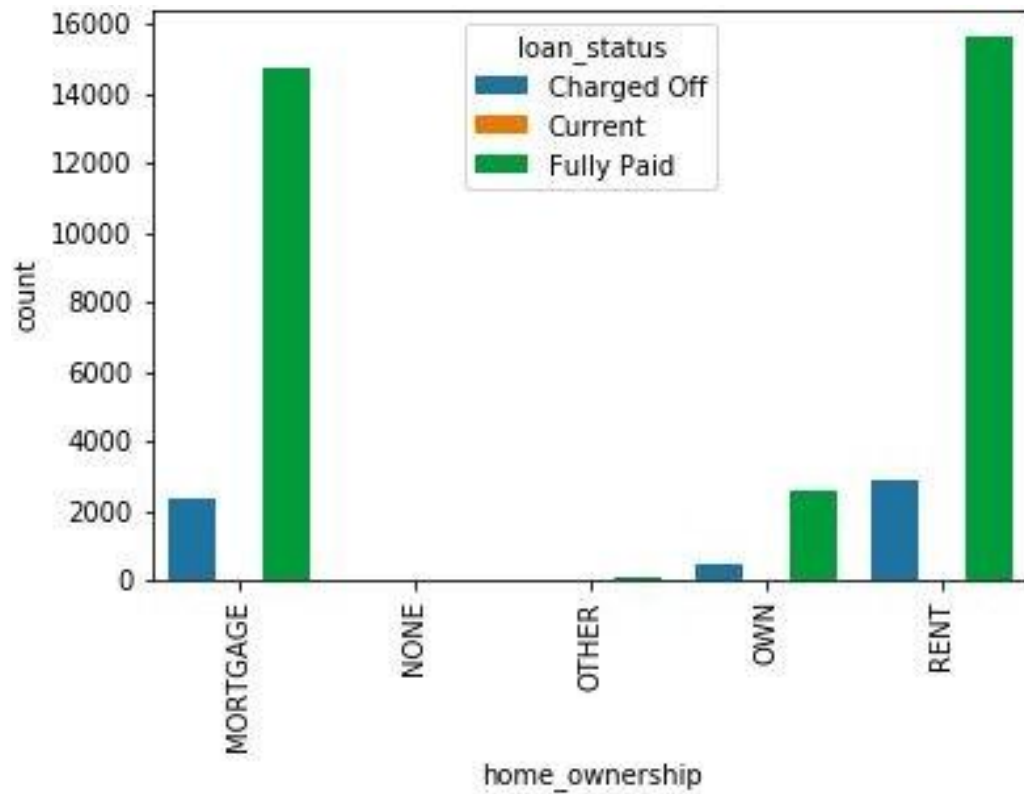


Observation:

1. Most of the loans granted were of 36 months.
2. And loans granted for 36 months have slightly high tendency to default.
3. Additionally, Debt Consolidation is the most common purpose and loans are granted for 36 months and 60 months.



Loan Status by Home Ownership



Observation:

Applicants having their home rented or on mortgage have almost equal tendency to default the loan.

Conclusion

Results

1. Loans that have higher interest rate have more defaulters. Check the background of applicant thoroughly if interest rate is high.
2. Extra interest rates or risk assessments must be done for the applicants belonging to CA state, as tendency to default is high.
3. When the purpose is debt consolidation then the lender must check the applicant thoroughly as it has high tendency to default (risky applicant).