Lending Club Case Study

Group Members

Surya Kiran Reddy Karri

Subhas Malik

Overview of Case Study

You work for a **consumer finance company** which specializes in lending various types of loans to urban customers. When the company receives a loan application, the company must decide 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 case study 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.

Problem Statement

Use EDA to understand how **consumer attributes** and **loan attributes** influence the tendency of default

Approach to the case study

- Data Sourcing
- Data Cleaning
- Univariate analysis
- Bivariate analysis
- Derived Metrics
- Visualizations

Data Cleaning

- Remove columns that have all null values
- Remove columns that have only 1 unique value
- Remove columns that have a majority of unique values
- Remove columns that don’t add any value to our analysis
- Filter rows to work with only non-Current loan statuses

Derived Metrics

We have derived couple of metrics:

Loan to Income Ratio = Loan Amount / Income

Percent Principal Received = Total Principal Paid / Loan Amount

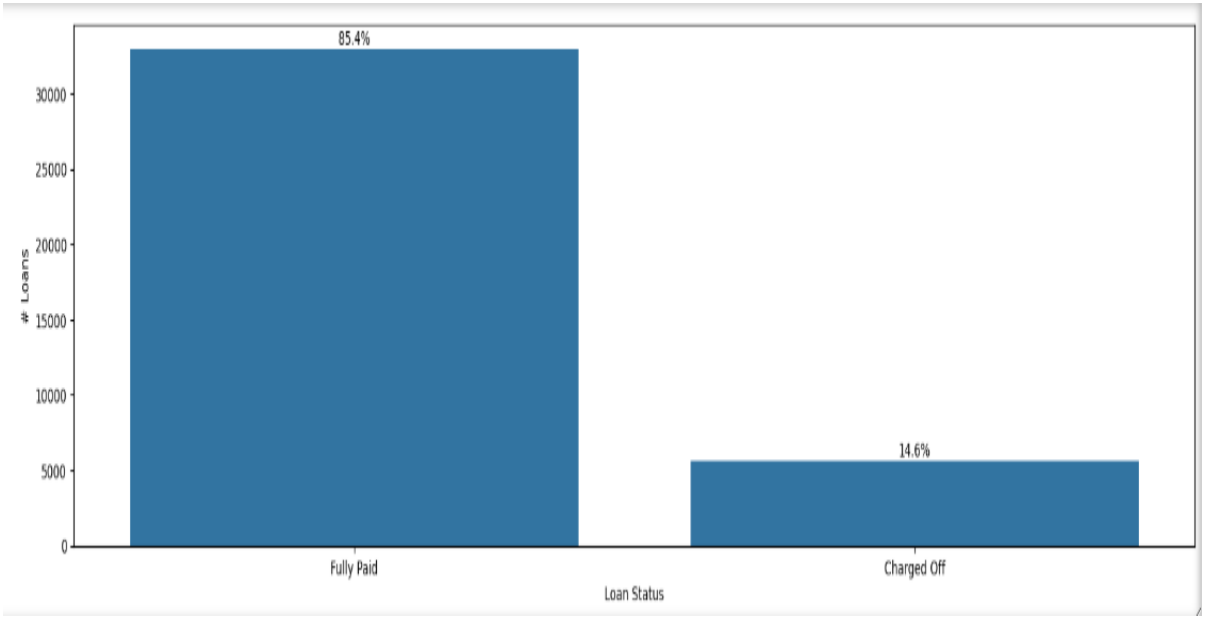
Both these metrics will be used to identify the incidence of default.

Data used for analysis

- After performing the initial checks and cleaning activities on the source data, we have arrived at 38577 rows of data.
- The following pages will describe some of the observations on this data.

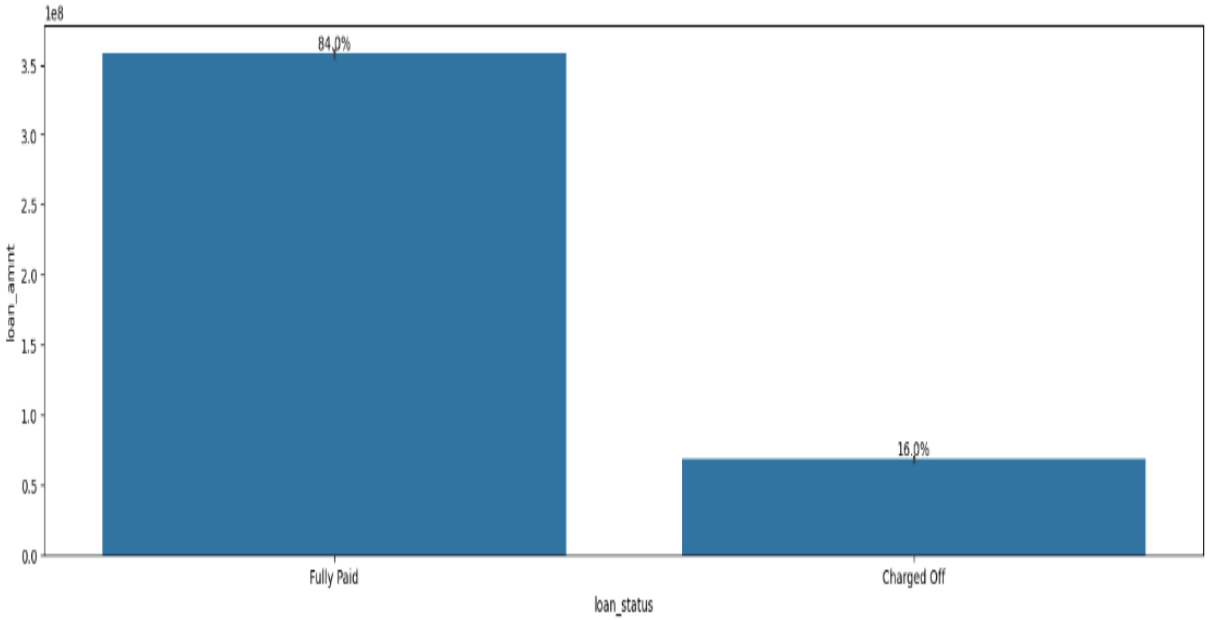
Observation 1

- ~15% of the loans have been charged off or defaulted.
- We will analyze further how we can identify these defaults so we can help the business save money.



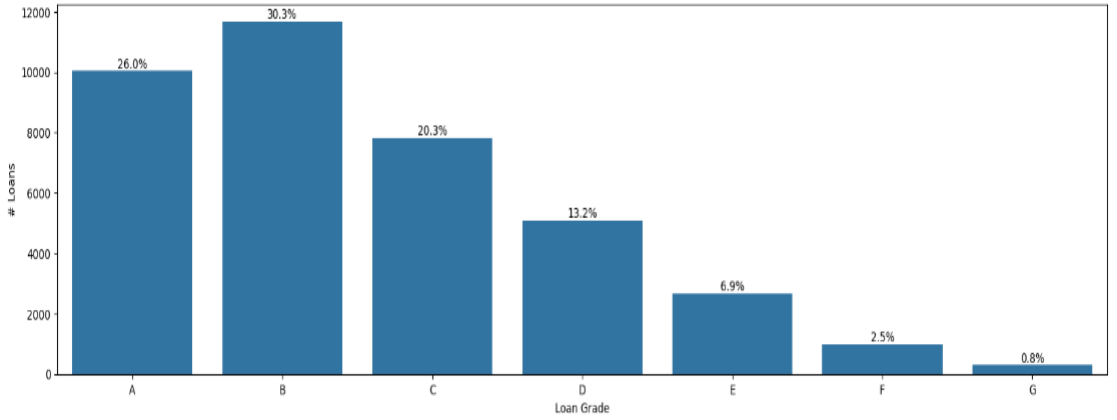
Observation 2

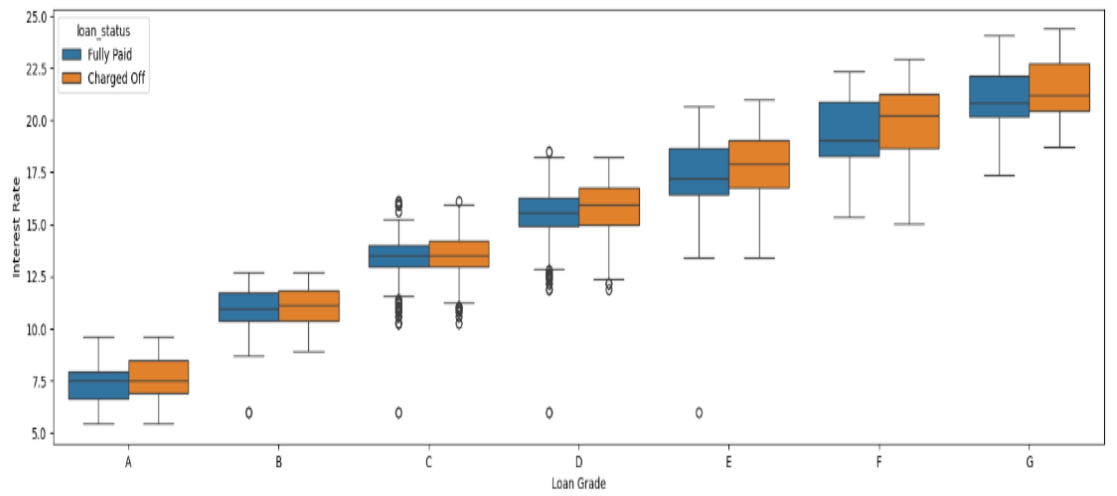
- Comparing this for the loan amounts, we notice that the total amount of loans that were defaulted amount to 16% of the total loan amount (fully paid + charged off)



Observation 3

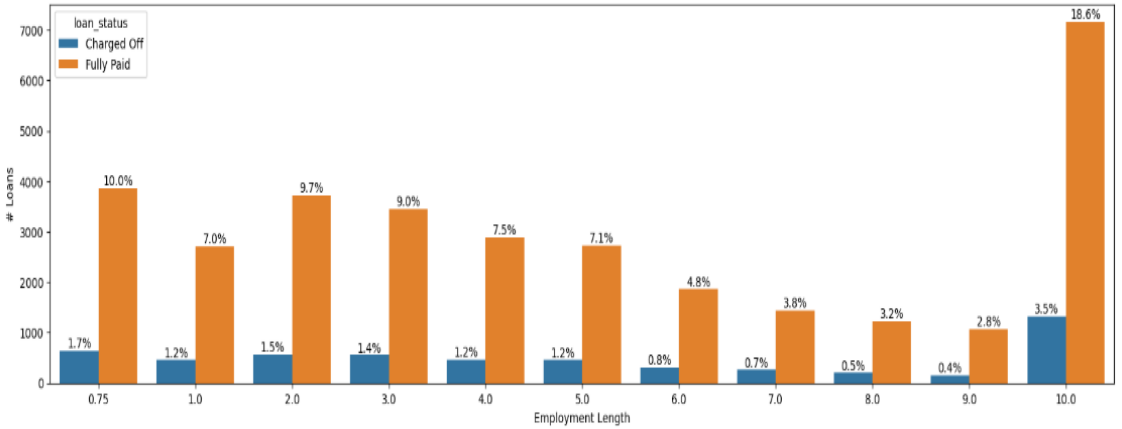
- The loan grades show that more than half the loans are high grade i.e. A or B (~56%)
- An interesting point to note is that ***mean value of the interest rates of default loans are higher compared to the fully paid ones in the same loan grade***.
- Moreover, higher the interest rate, more likely is it that the loan would be defaulted (across all loan grades).

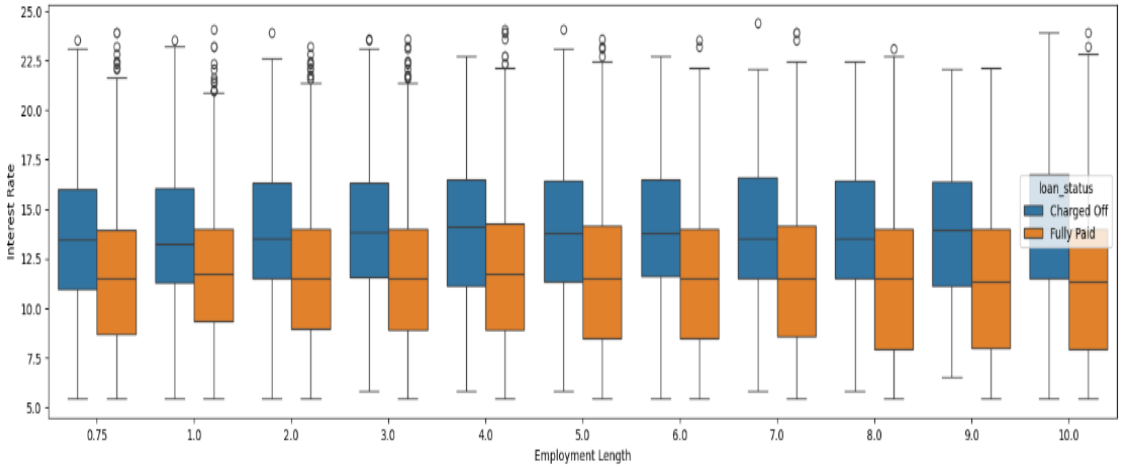




Observation 4

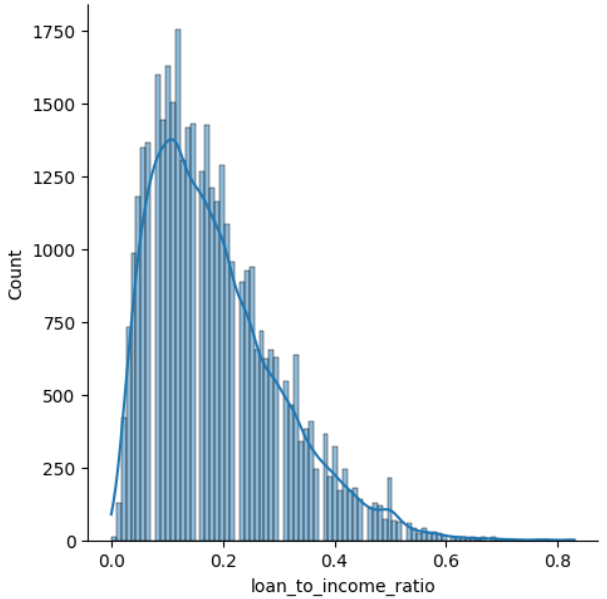
- We can notice that most loans are availed by people with more than 10 years of experience and people with less than 1 year of experience (~34%)
- When we compare the employment length with the interest rates and check for the loan status, we notice that the default loans all have higher interest rates. The ***means of the default loans are closer to the 75 percentile of the fully paid loans***.
- **A higher interest rate could be one deterrent for paying the loan.**

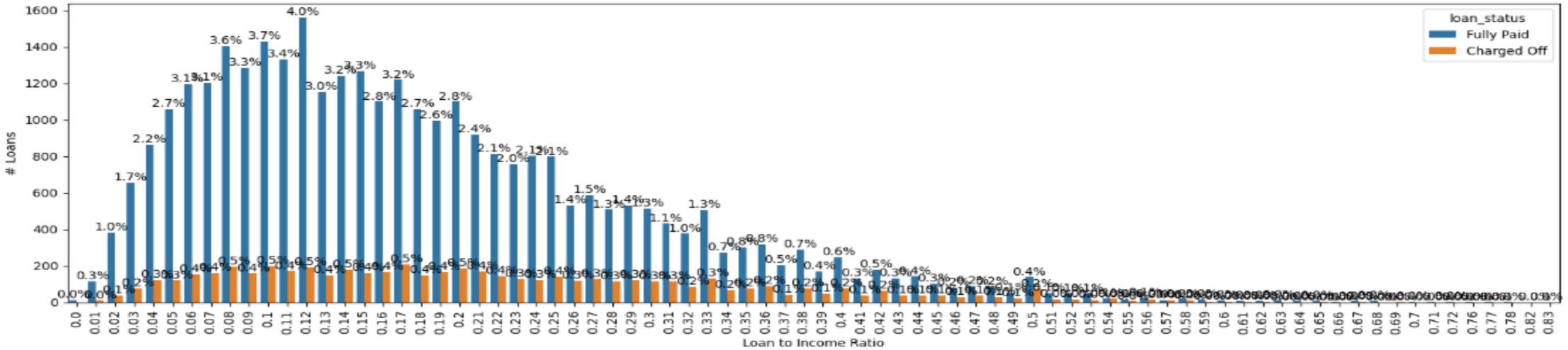




Observation 5

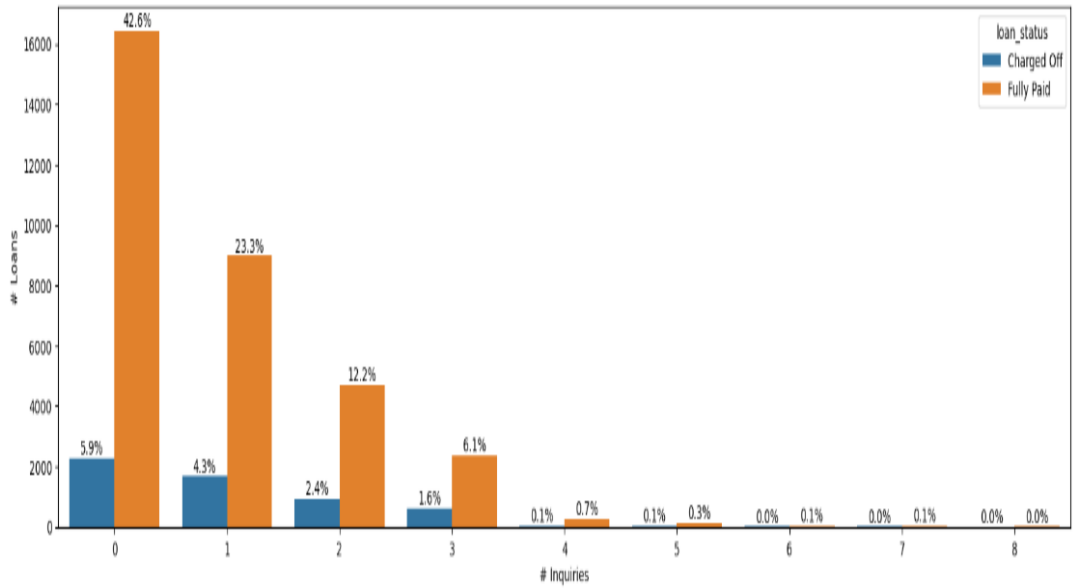
- The loan to income ratio distribution is left skewed i.e. most loans are availed for lower ratios.
- Another interesting point to note here is that there is a sweet spot ***ratio of around 0.12 where the % of defaults are lesser.***





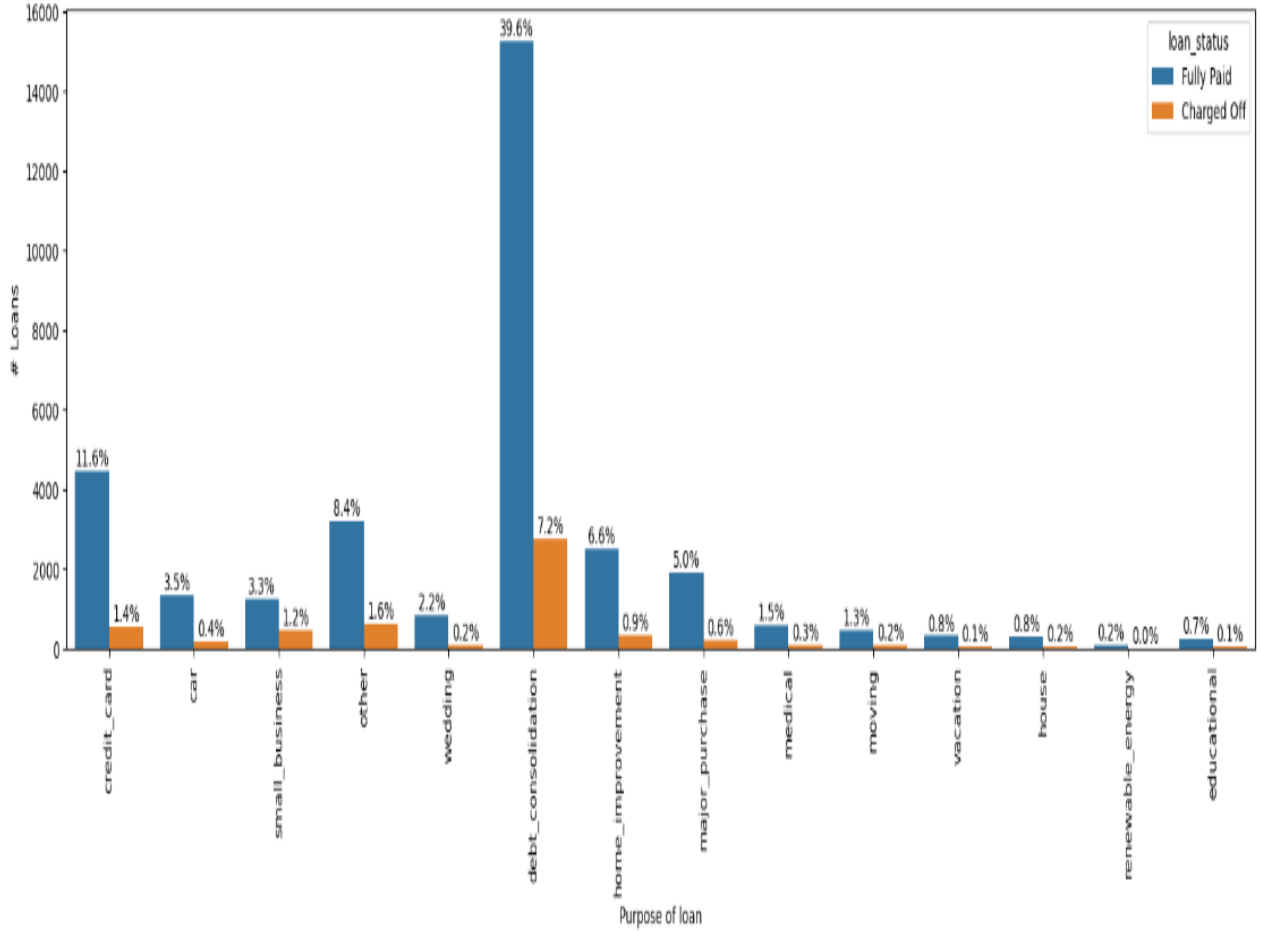
Observation 6

- Another interesting observation is regarding the # of inquiries in last 6 months. As the inquiries increase the incidence of default will increase. We can see that while the number of loans are decreasing the ***% of defaults compared to the total loans is increasing with the number of inquiries***.



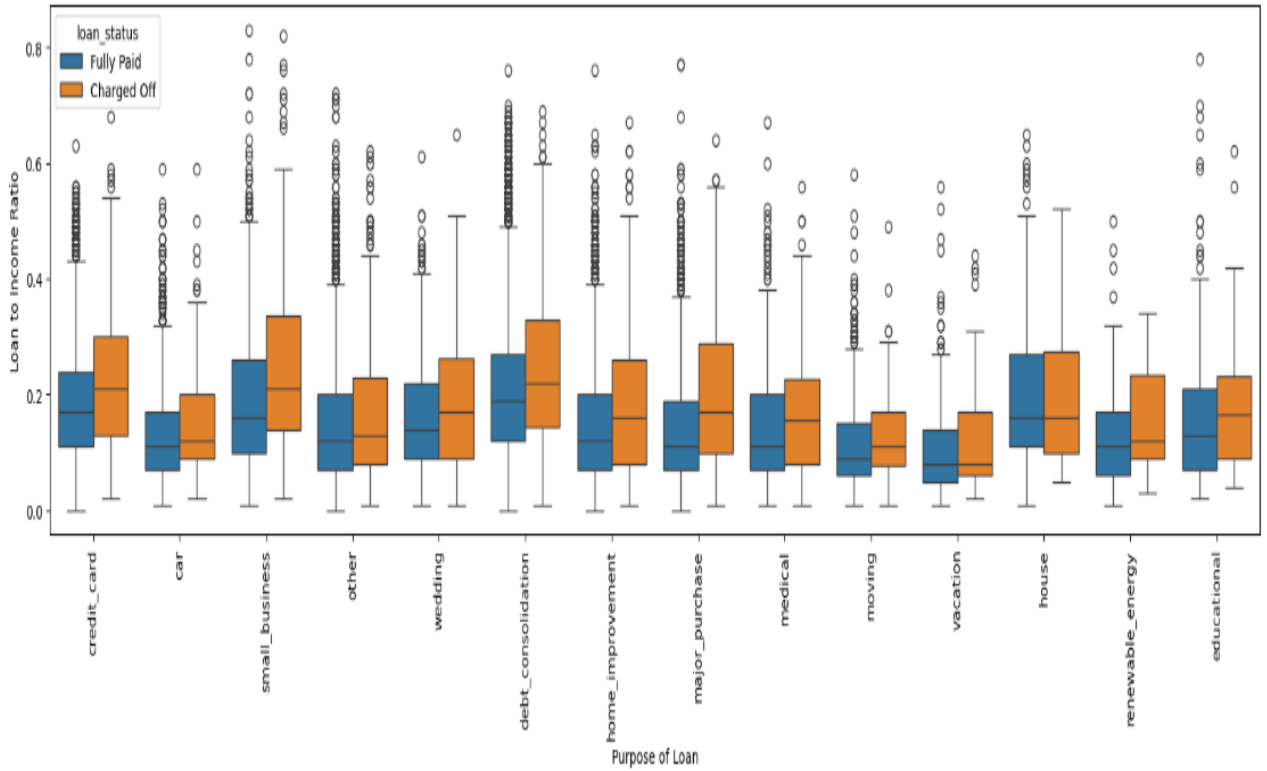
Observation 7

- Majority of loans have been taken for debt reconciliation and credit card.
- ***Loans taken for purpose of wedding and cars tend to have the least number of defaults as a percentage of the total loans taken for that purpose.***



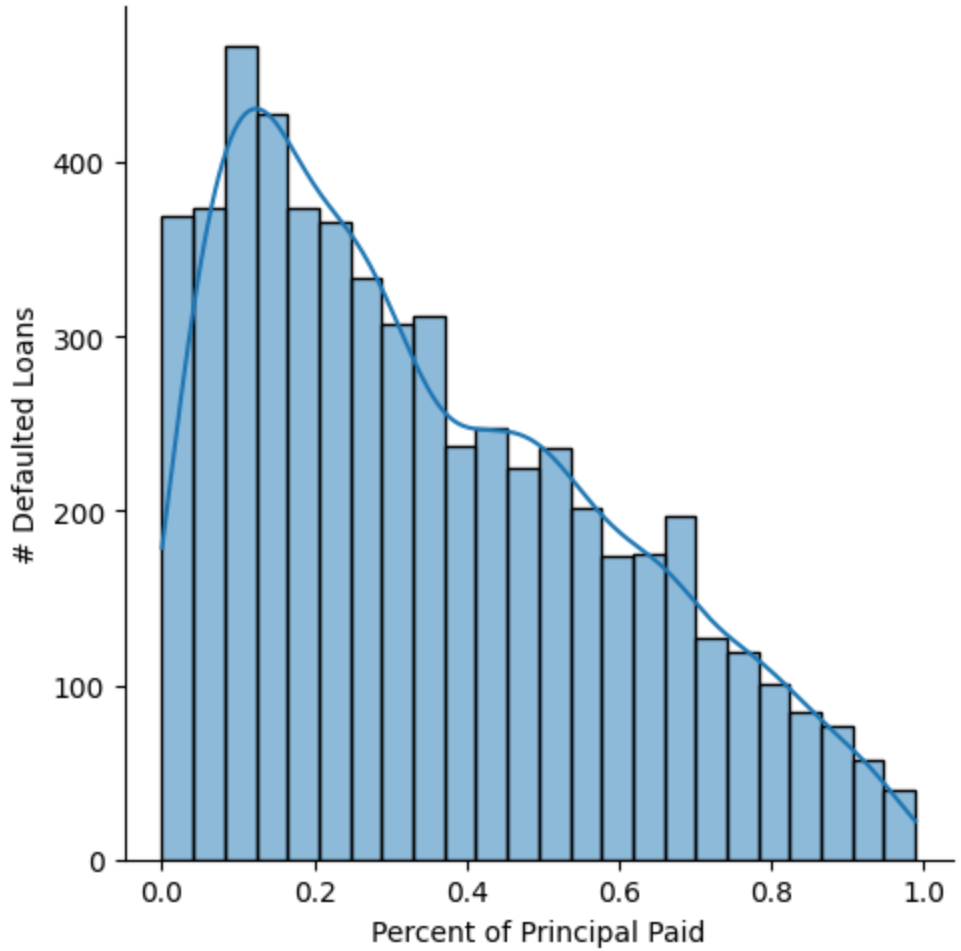
Observation 8

- Loan to income ratio also plays a crucial part in identifying the default incidence.
- Loans with higher loan to income ratio tend to be defaulted.



Observation 9

- For default loans, the distribution of percentage of principal paid versus to loan amount is left skewed and we can notice that ***most loans have been defaulted after paying less than 20% of the loan amount.***



Observation 10

- Most loans have been taken by people from CA state
- ***The default ratio is less for TX.***