

LENDING CLUB CASE STUDY

GROUP MEMBERS:

- ▶ Sachin Rajput
- ▶ Pragati Singh



About The Company

- ▶ 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.

OBJECTIVE

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.

PROBLEM STATEMENT

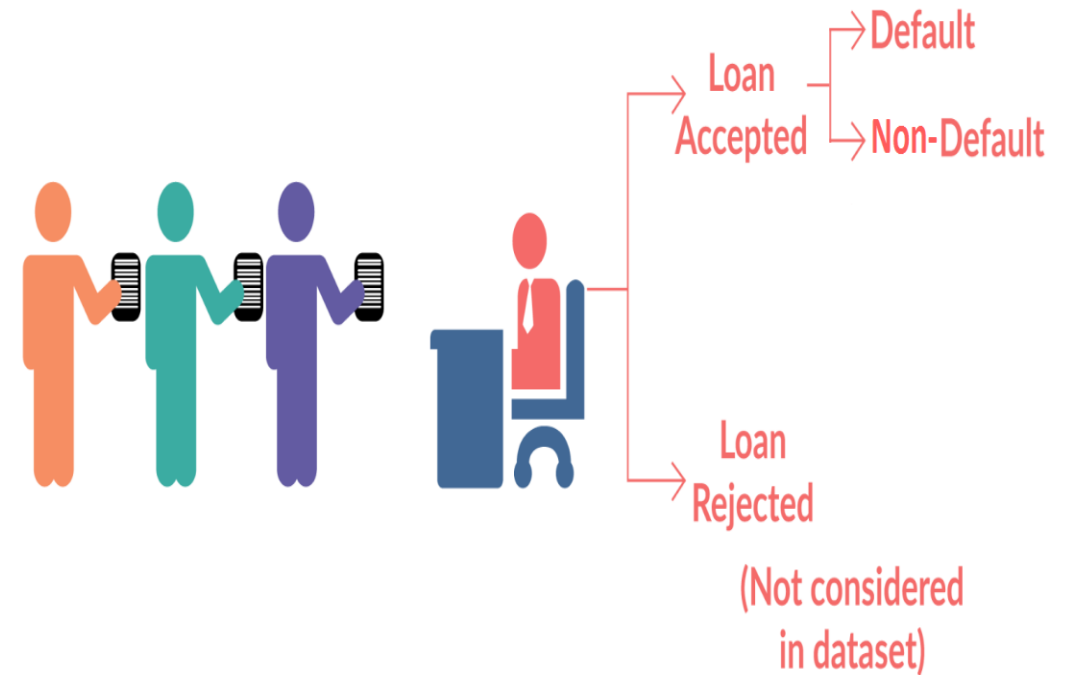
The data contains the information about past loan applicants and whether they 'defaulted' or not. The aim is to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

In this case study, you will use EDA to understand how **consumer attributes** and **loan attributes** influence the tendency of default.

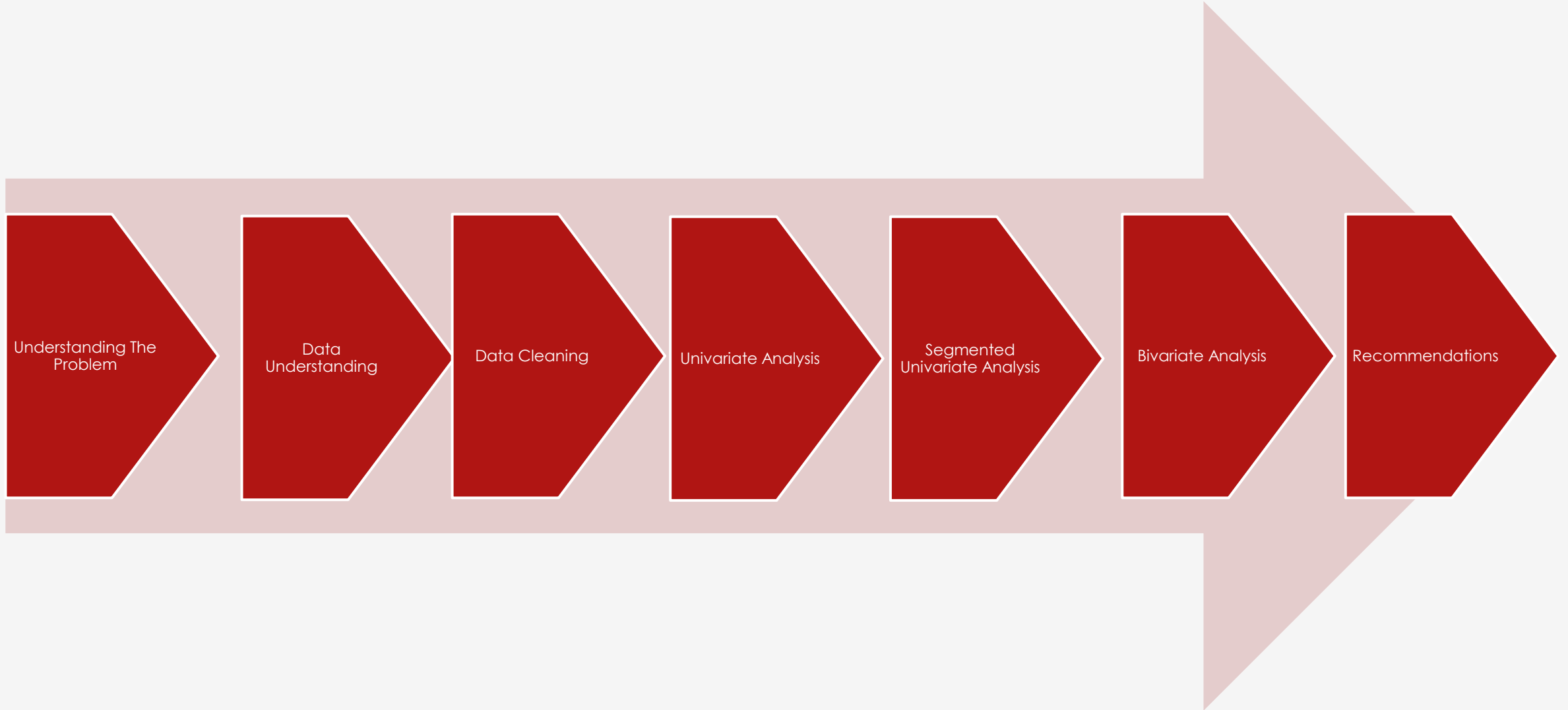
Description

1. **Loan accepted:** If the company approves the loan, there are 3 possible scenarios described below:
 1. **Fully paid:** Applicant has fully paid the loan (the principal and the interest rate)
 2. **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'.
 3. **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
2. **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)

LOAN DATASET



Analysis Approach

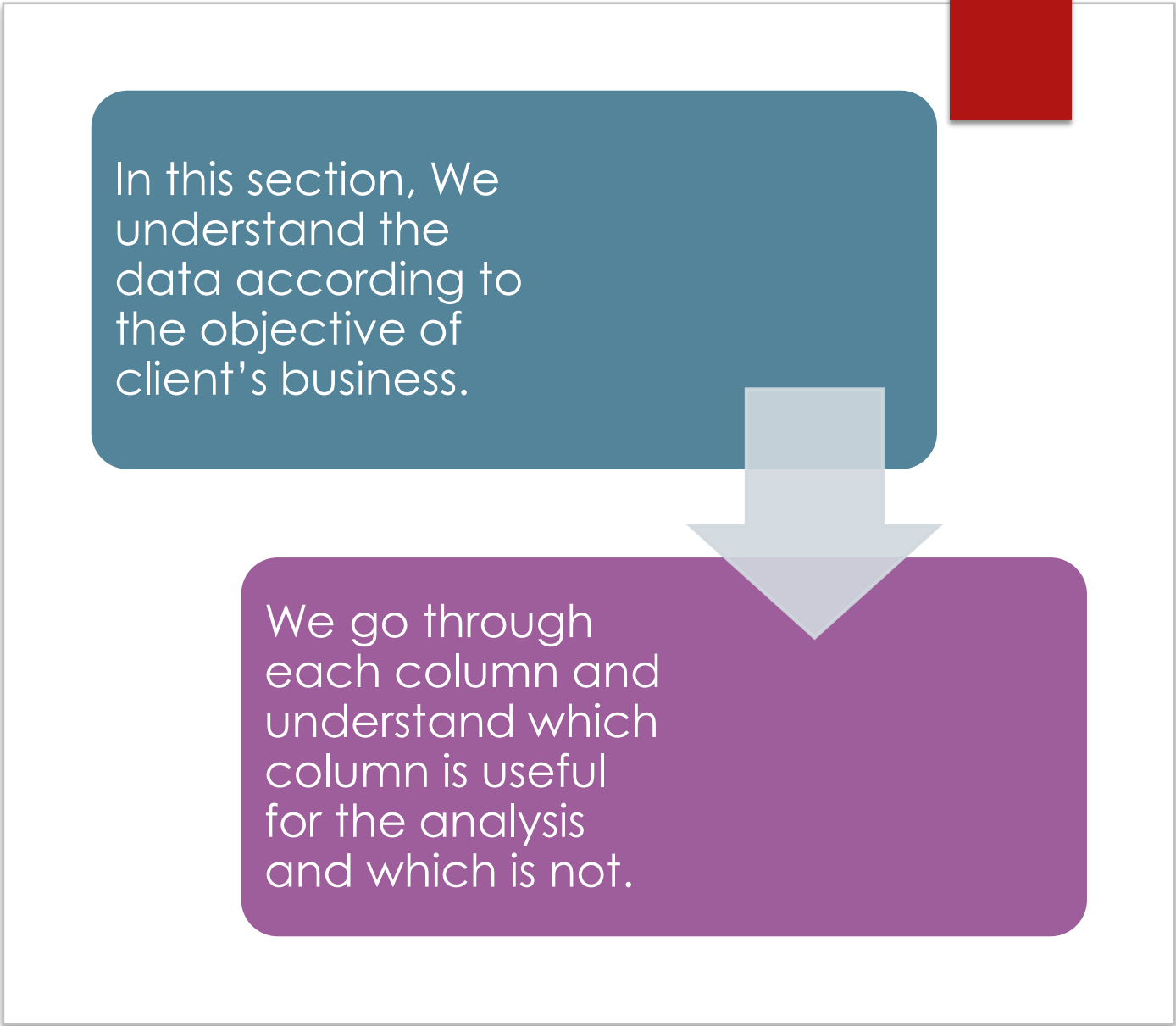


Understanding The Problem

The first step of any analysis is understanding the client's problem and know what are their requirements. Basically, the objective of the business.



Data Understanding



In this section, We understand the data according to the objective of client's business.

We go through each column and understand which column is useful for the analysis and which is not.

DATA CLEANING

Dropping columns with null values,
all random values or single category
value

or behavioural variables.

Converting values to numeric form
(proper Int and Float type).

Univariate Analysis

Analysing each column and plotting the distribution of each column.

Check distributions and frequencies of various numerical and categorical variables

Segmented Univariate Analysis

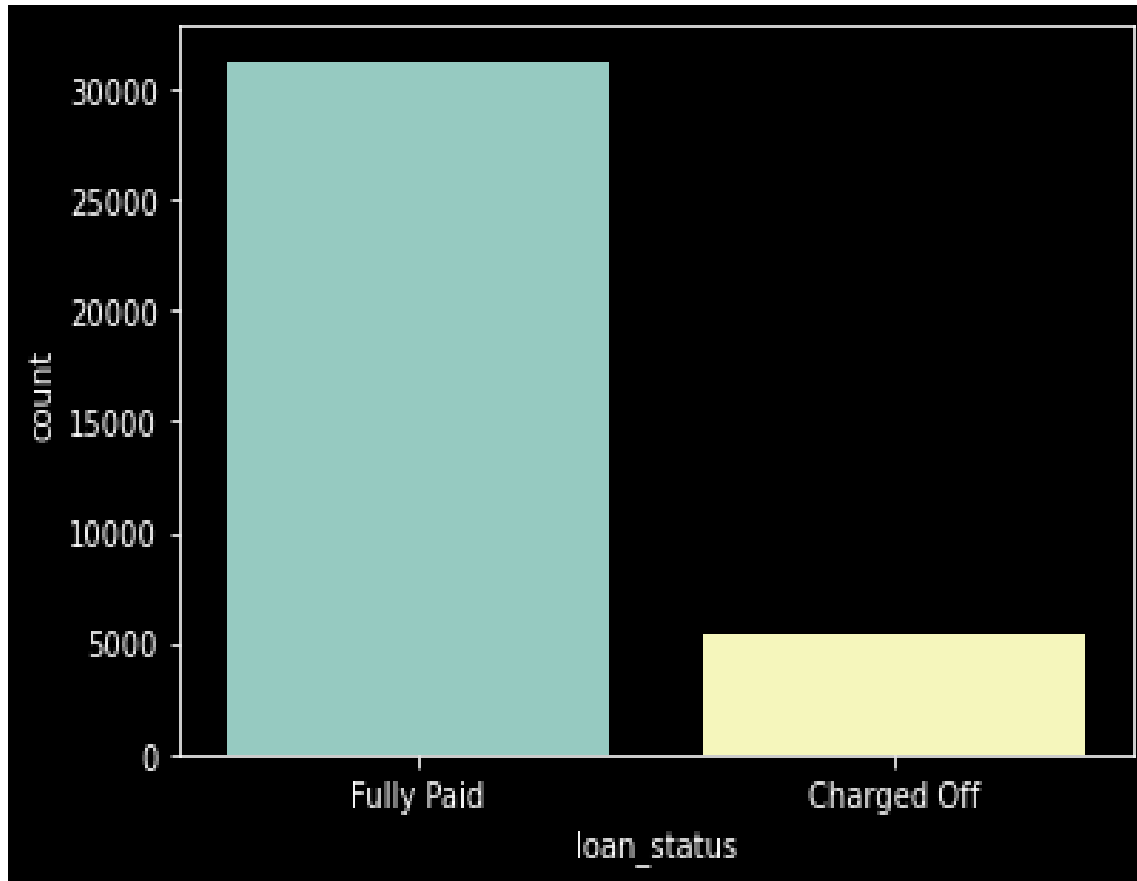
Analyse variables against segments of other variables

Data Variable is analysed in subsets and is very useful as it can show the change metric in pattern across the different segments of the same variable

Bivariate Analysis

Bivariate Analysis is one of the simplest form of quantitative analysis.

It involves analysis of two variables and do correlation analysis and check how two variables affect each other or a third variable.

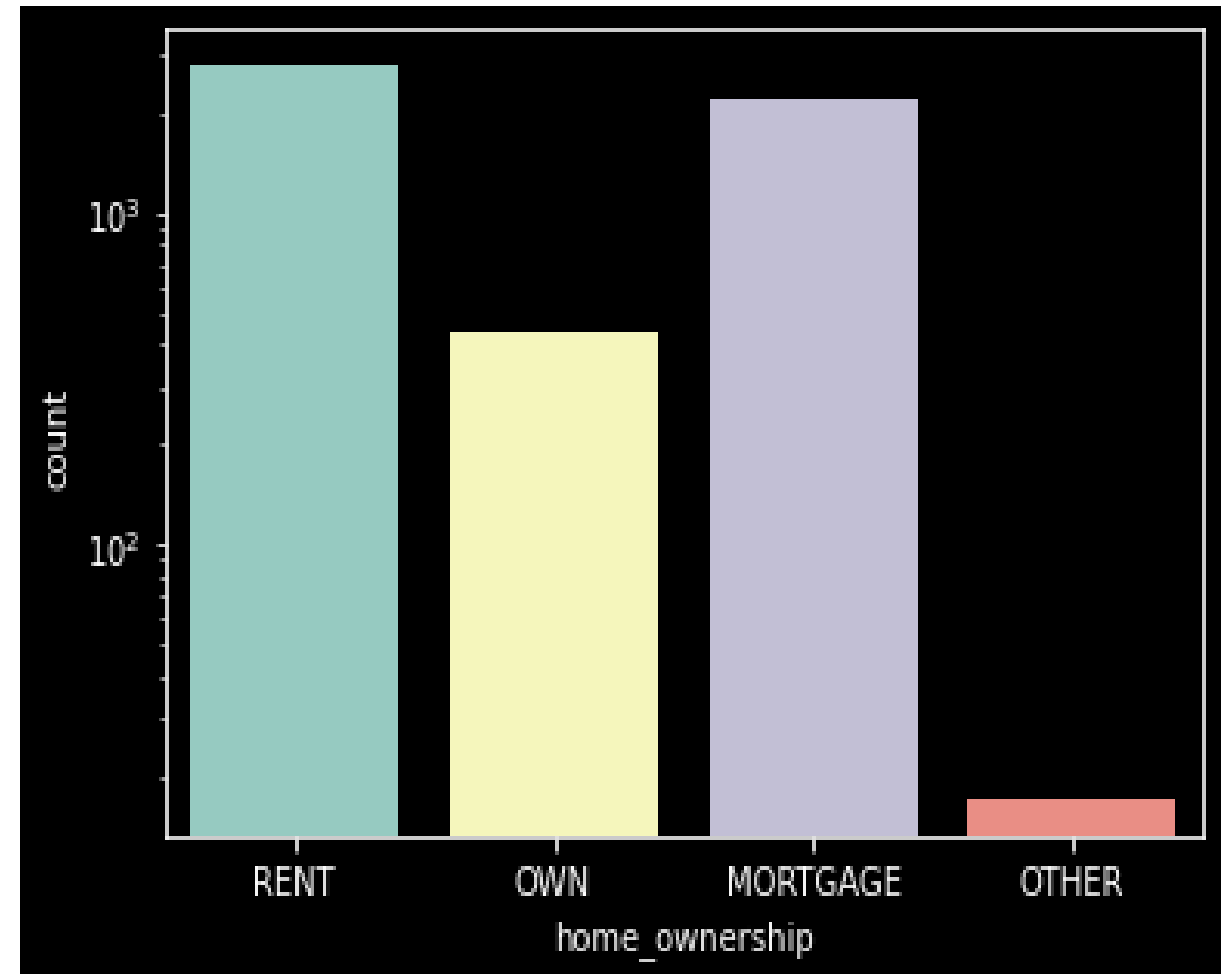


LOAN STATUS ANALYSIS

BY THIS VISUALIZATION WE CAN COUNT THE OBSERVATIONS OF LOAN STATUS FOR FULLY PAID AND CHARGED OFF.

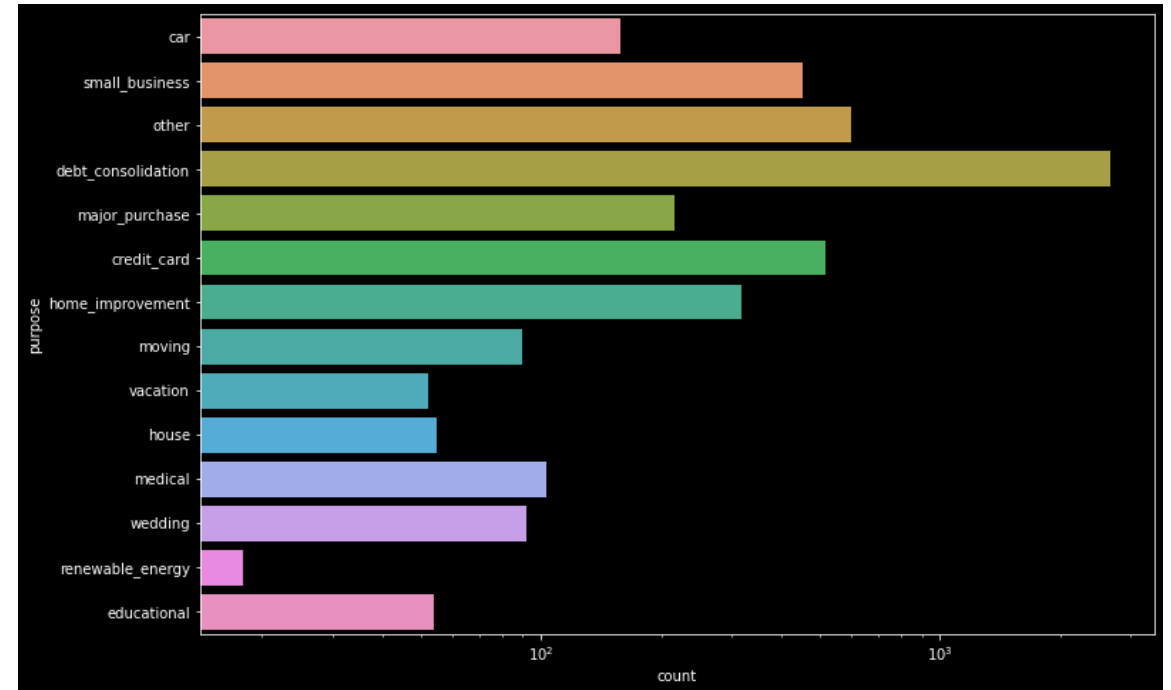
Relationship b/w Home Ownership & Charged Off Loan Status

Rent and Mortgage Home ownership are more likely to charged off as compared to OWN and Other.



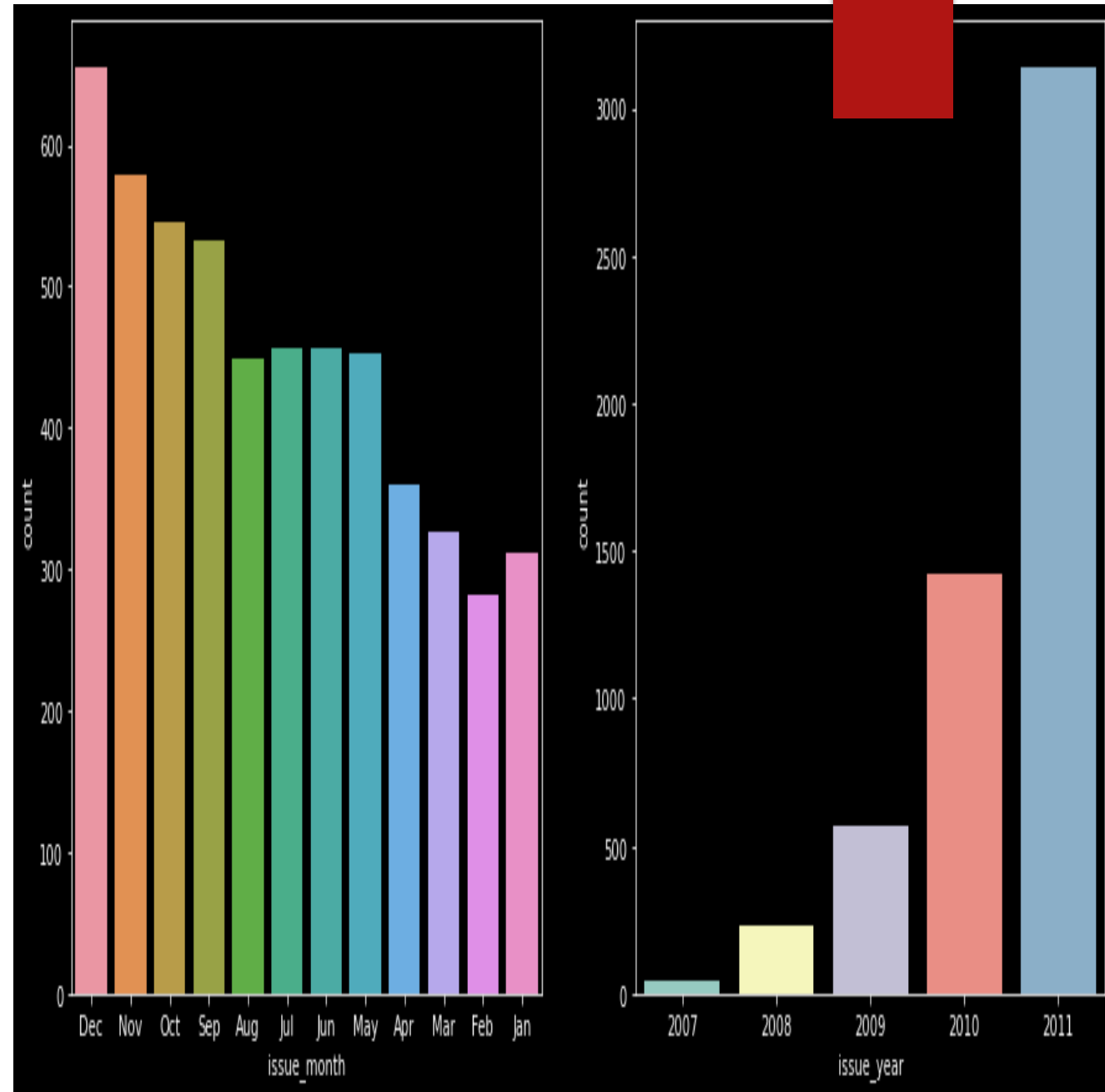
Relationship b/w the purpose and Charged off Loan status

By this we can see that maximum number of loans are from Debt Consolidation and minimum number of loans are from renewable energy whose loan status is Charged Off



Relation between Defaulters and Loan Funded Month and Year

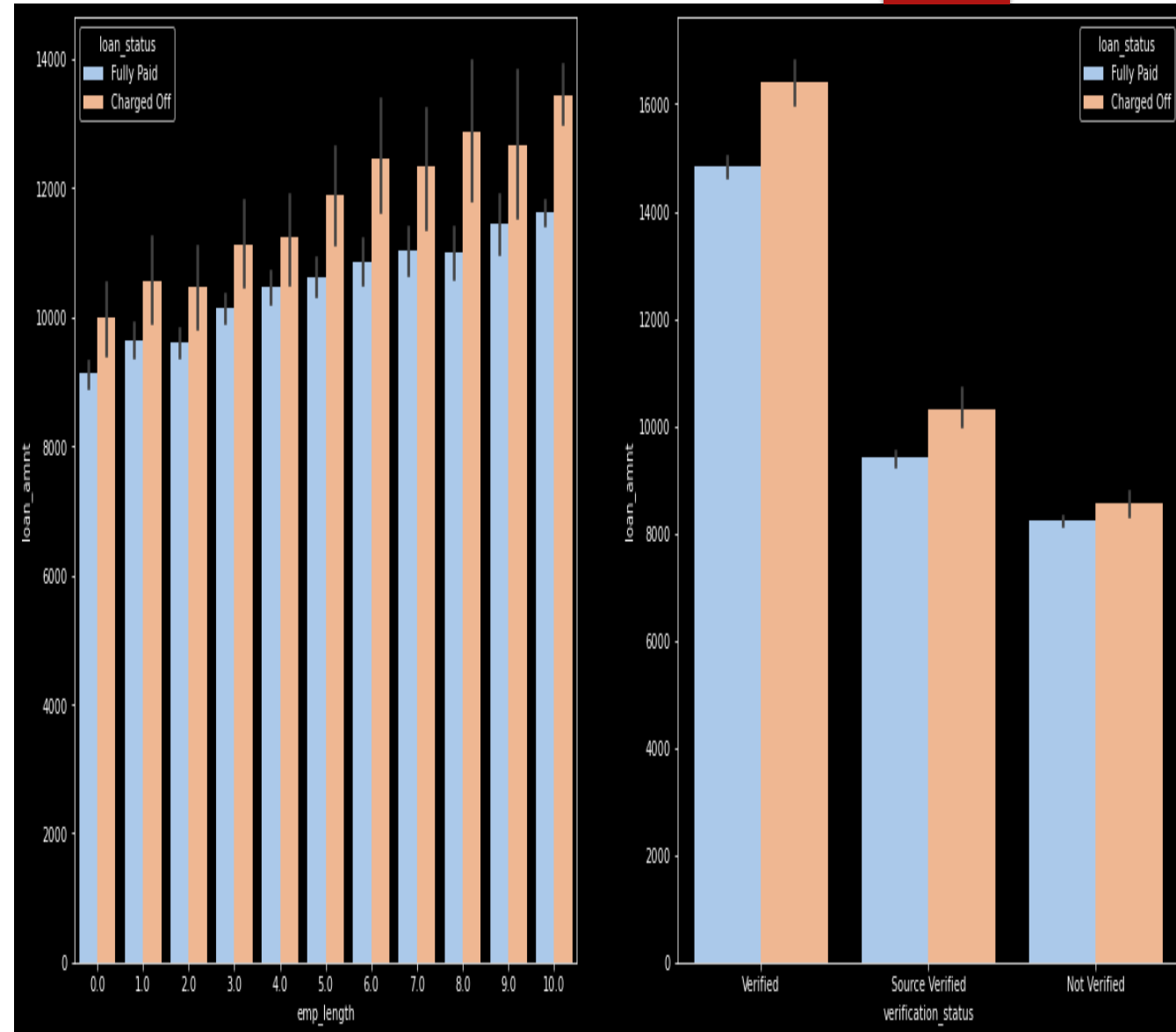
Loan issued in the Month of December have more defaulters and year by year charged off is increasing.



Employment years, loan amount, verification status and loan status

Employees with longer working history got the loan approved for a higher amount.

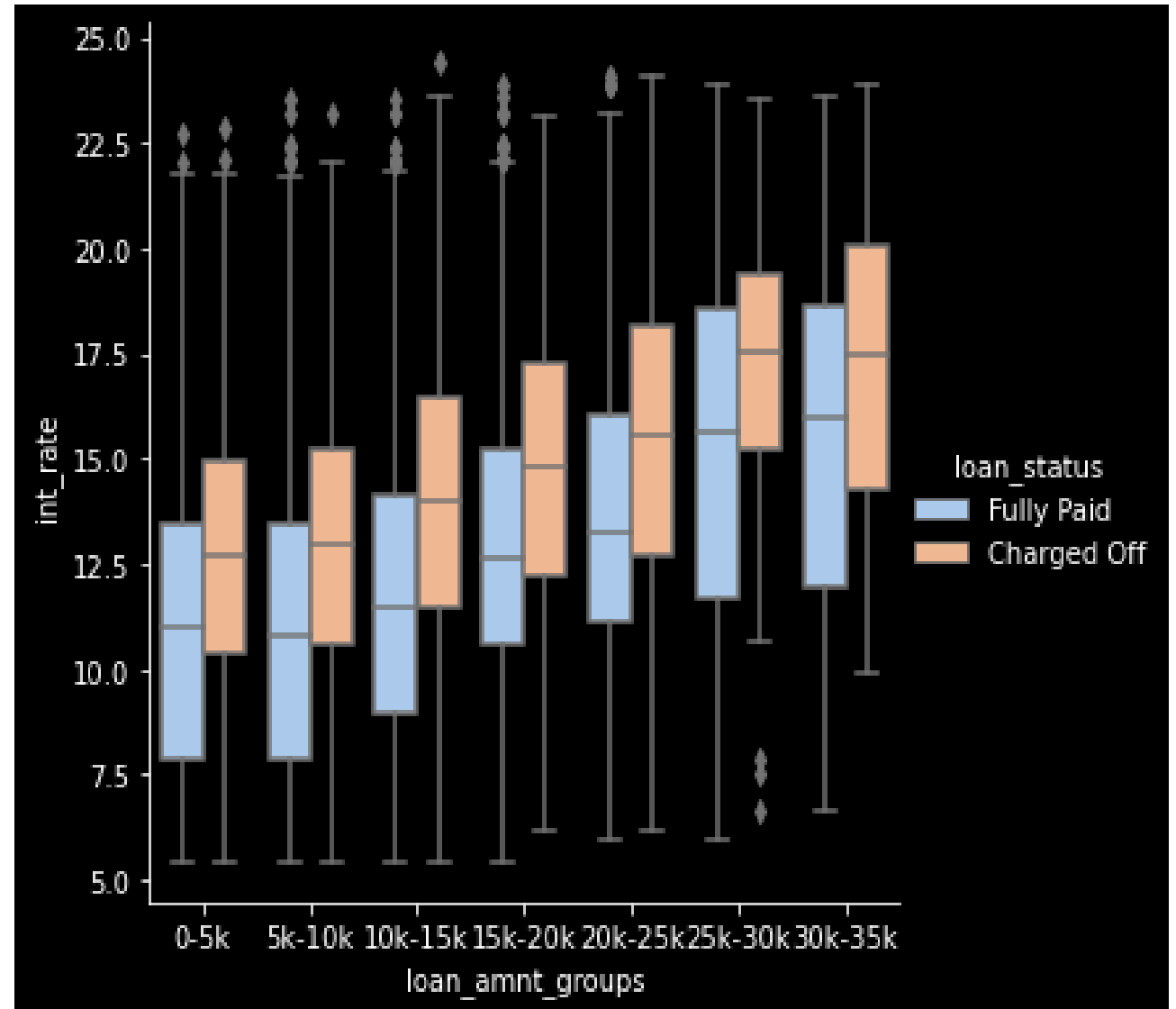
- Looking at the verification status data, verified loan applications tend to have higher loan amount. Which might indicate that the firms are first verifying the loans with higher values.



Analyzing Interest Rate, Loan Amount Groups and Loan Status

The interest rate for charged off loans is pretty high than that of fully paid loans in all the loan amount groups.

- This can be a pretty strong driving factor for loan defaulting.



CONCLUSION

- ▶ **The above analysis with respect to the charged off loans. There is a more probability of defaulting when :**
 - Applicants taking loan for 'home improvement' and have income of 60k -70k
 - ▶ - Applicants whose home ownership is 'MORTGAGE and have income of 60-70k
 - ▶ - Applicants who receive interest at the rate of 21-24% and have an income of 70k-80k
 - ▶ - Applicants who have taken a loan in the range 30k - 35k and are charged interest rate of 15-17.5 %
 - ▶ - Applicants who have taken a loan for small business and the loan amount is greater than 14k
 - ▶ - Applicants whose home ownership is 'MORTGAGE and have loan of 14-16k
 - ▶ - When grade is F and loan amount is between 15k-20k
 - ▶ - When employment length is 10yrs and loan amount is 12k-14k
 - ▶ - When the loan is verified and loan amount is above 16k
 - ▶ - For grade G and interest rate above 20%