Lending Club Case Study

Author: Vinit Kumar

Abstract

- Lending club is the largest online loan marketplace, facilitating personal loans, business loans, financing of medical procedures and various other loan forms.
- Lending club matches borrowers who are seeking a loan with investors looking to lend money and make return.
- When company receives loan application company need to make decision for loan approval based on applicant profile to safeguard credit loss.
- Our objective is to identify such factors which causes loan default so that Lending club can can make loan approval efficient and reduce credit loss.

Problem Statement:

- For the borrower's loan application, 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.
 - -If the applicant is not likely to repay the loan, then approving the loan may lead to a financial loss
- If one is able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. Identification of such applicants using EDA is the aim.

Objective:

• Lending club wants to understand the driving factors behind loan default. So that lending club can utilize this knowledge for its portfolio and risk assessment. We need to Provide the EDA based analysis on driving factors and possible recommendations.

Problem solving methodology

Data Analysis Data Analysis Data Analysis

Data Cleaning

Removing the null valued columns, unnecessary variables and checking the null value percentage and removing the respective rows.

Data Understanding

Working with the Data Dictionary and getting knowledge of all the columns and their domain specific uses

Univariate Analysis

Analysing each column, plotting the distributions of each column.

Segmented Univariate Analysis

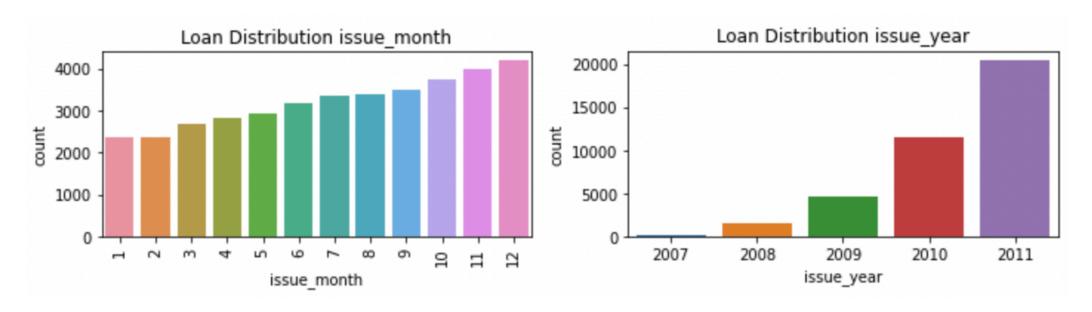
Analysing the continuous data columns with respect to the categorical column

Bivariate Analysis

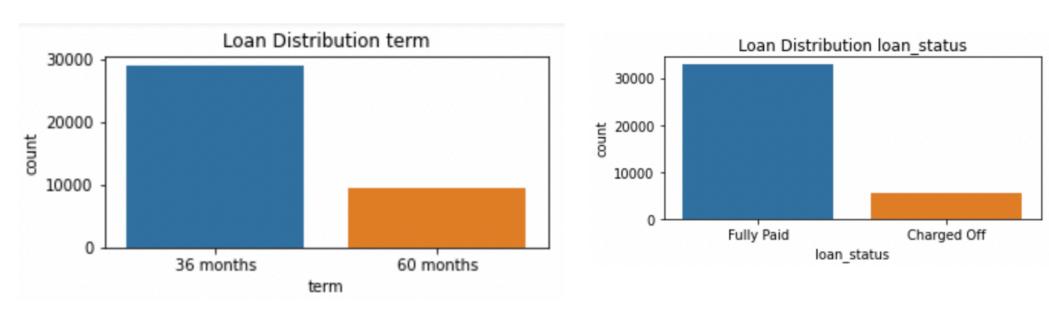
Analysing the two variable behaviour like term and loan status with respect to loan amount.

Recommendations

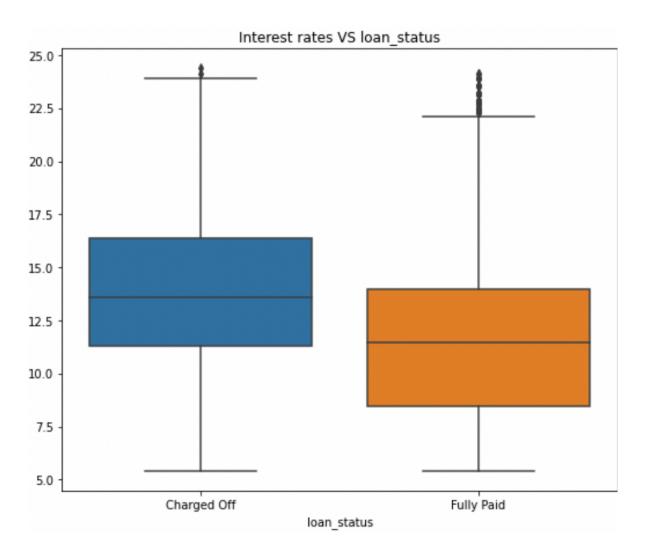
Analysing all plots and recommendations for reducing the loss of business by detecting columns best which contribute to loan defaulters.



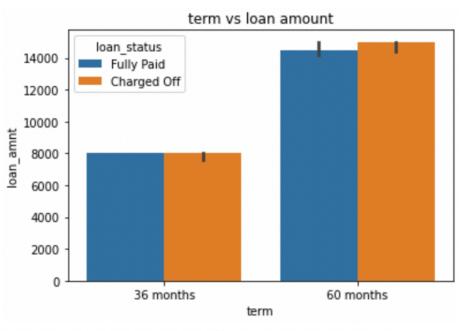
- Lending club has expanded loan approval almost 100% and doubled loan approval count every year.
- Also the issued month of loans is also increasing from January to December. In the final quarter of year there are more loans issued this could be because of vacation and christmas.

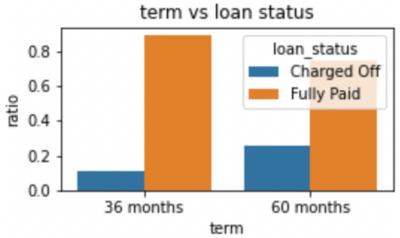


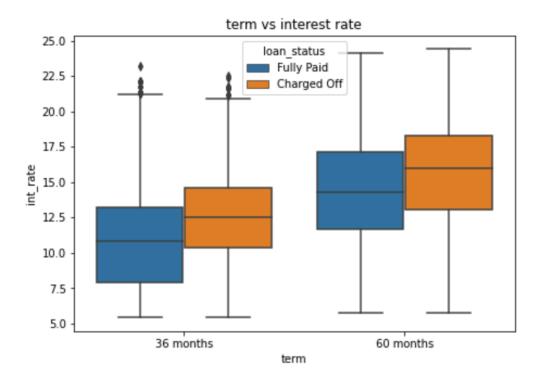
- There are only two loan terms 36 and 60 months. Around 75% borrowers took loans with 36 months term.
- The charged off borrowers are around 15% and fully paid is around 85% in the given data set.



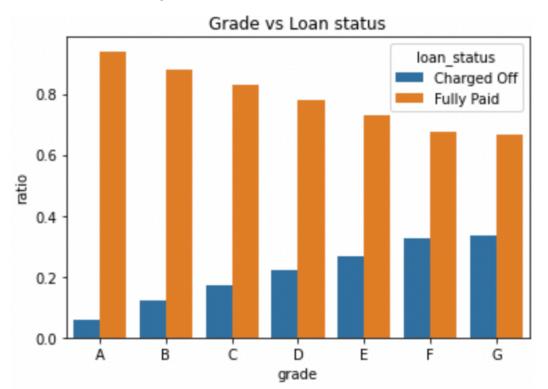
When the loan interest rate is high there is high chance of loan getting defaulted.

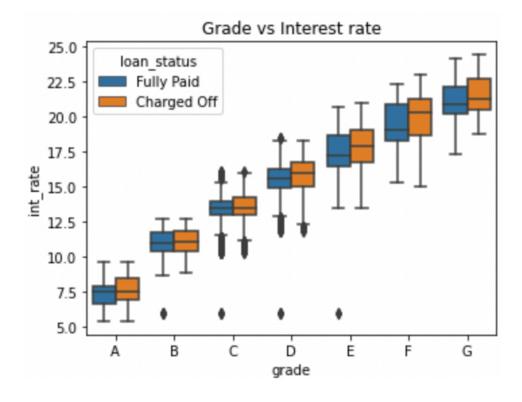




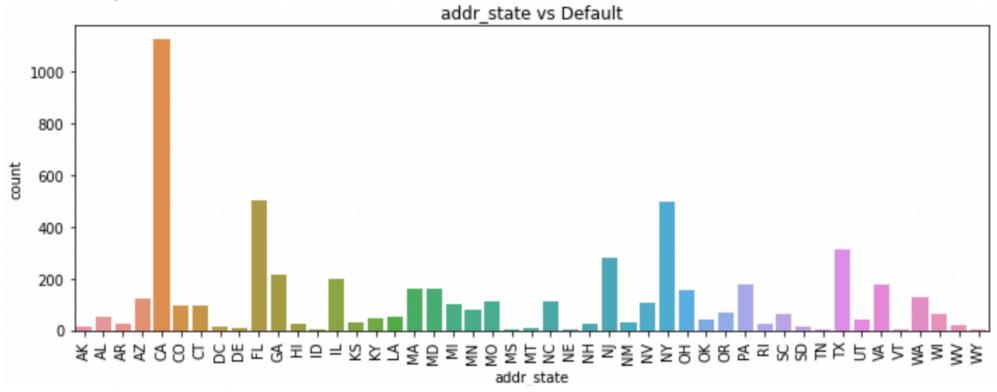


The default rate is high in 60 months tenure because most people took high loan amount with high interest rate in it and they faced difficulties in returning the sum to bank.

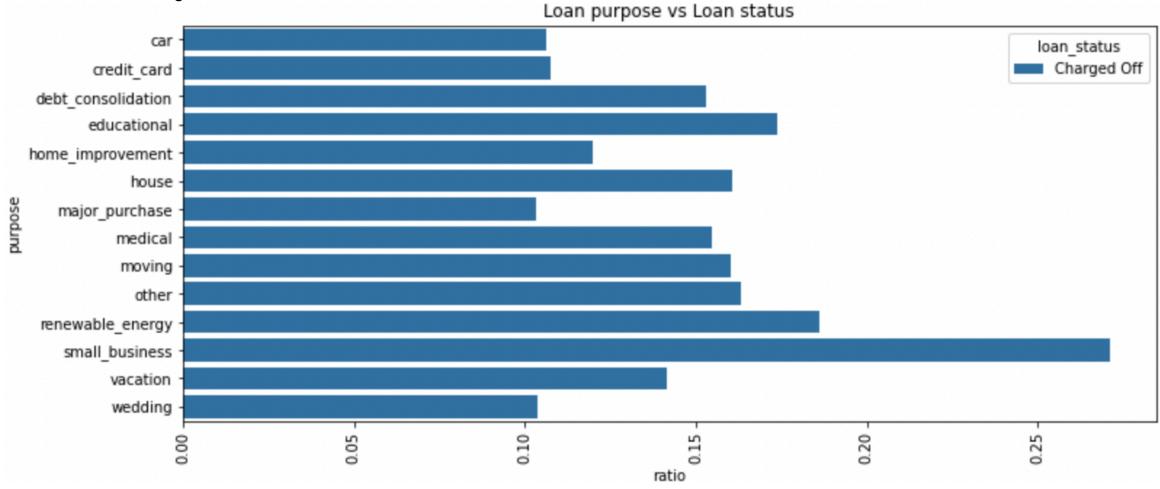




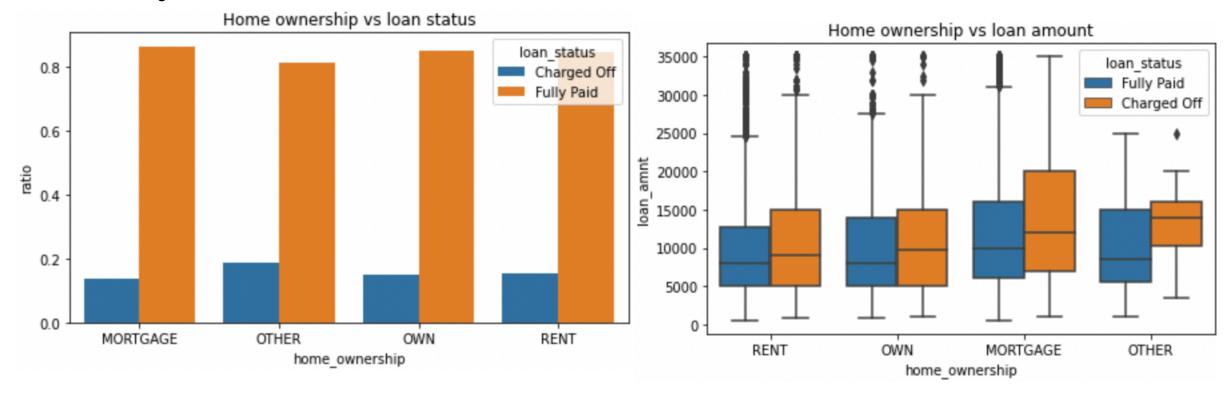
- Grades are very good category to tell the borrower probability of defaulting the loan.
- The Lower grades(E,F,G) have higher chances of defaulting the loan than Higher ones(A,B)
- Also the Lower grades are getting loans for higher interest rates which might be the cause for loan default.



The loans which are given for CA, FL and NY state borrower's are defaulted more than other states.



Borrower's who took loans for small business purpose have defaulted more.



- There is around 20% chance of loan default in each home ownership category.
- From the 2nd plot we can see the people with higher loan amounts in **mortgage** home ownership has high default rate than others.

Other Analysis Observations:

There is probability of applicant defaulting loan under below scenarios:

- If interest rate and DTI is higher than loan for tenure 60 months.
- Grade E,F,G tend to be defaulting with high loan amount >14K.
- Higher loan amount to employee with service length more than 10 years
- Small business and house loan with higher loan amount greater than 14K on high interest rate.
- Applicants whose home ownership is Mortgage and have loan greater than 14K
- Applicant with service length more than 10 years and have loan amount greater than 12K
- When annual income is verified and loan is greater than 14K and interest rate greater than 12%
- When grade is G and interest rate greater than 20%
- Instalment above 300 for the tenure of 60 months.

Conclusion & Recommendations

- Lending club should reduce the high interest loans for 60 months tenure, they are prone to loan default.
- Grades are good metric for detecting defaulters. Lending club should examine more information from borrowers before issuing loans to High grade (A->G).
- Lending Club should control their number of loan issues to borrowers who are from CA, FL and NY to make profits.
- Small business loans are defaulted more. Lending club should re-evaluate strategy on loan amount and Interest rates for this loan type.
- Borrowers with mortgage home ownership are taking higher loans and defaulting the approved loans. Lending club should stop giving loans to this category when loan amount requested is more than 14000.