Lending Club Case Study

Group Members:

Amit Singh

Sohini Banerjee

Overview

- **Lending Club is a 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,** that is, he/she is likely to default, then approving the loan may lead to a **financial loss** for the company

- The objective of this case study 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.

Assumptions

a. In the dataset, there are some customer behavior variables which are generated after the loan is approved and hence not available at the time of loan application. Thus, these variables cannot be used as the predictors for credit approval.
   - delinq_2yrs,
   - earliest_cr_line,
   - inq_last_6mths,
   - open_acc,
   - pub_rec,
   - revol_bal,
   - revol_util,
   - total_acc,out_
   - prncp, out_
   - prncp_inv,
   - total_pymnt,
   - total_pymnt_inv,
   - total_rec_prncp,
   - total_rec_int,
   - total_rec_late_fee,
   - recoveries,
   - collection_recovery_fee,
   - last_pymnt_d,
   - last_pymnt_amt,
   - last_credit_pull_d,
   - application_type

Assumptions (Contd..)

a. The records for which the "loan_status" is marked as "current" are neither fully paid nor defaulted. Hence, these records have been removed before analyzing the dataset further.

b. Columns / features for which >30% of the values are missing have been discarded as the percentage of missing values are quite high.

c. The columns "emp_title" & "title" contain text values which will not impart any additional value to the analysis.

d. The columns "id","url" and "member_id" all have unique values. Hence, we will use "id" as the primary key and remove the other 2 columns.

Univariate Analysis

a. **Loan Amount:** From the box plots, it is observed that a greater number of people took loan amount of 10000, and also median of distribution is 10000. Very few people took more than 30000 loan amount.

b. **Funded Amount:** Funded amount data shows similar trends as that of the loan amount. This indicates that the Lending Club had approved most of the applied loan amount.

c. **Interest Rate:** From the boxplot, it is observed that most of the interest rates lie between 9% to 14.5%. Some people took loan at a higher interest rate of 22.5%.

d. **Term:** The bar graph shows that more borrowers have taken loans with 36 months tenure rather than 60 months tenure.

e. **Loan Status:** The bar graph shows that around 85% of the borrowers have fully paid the loan while around 14% have defaulted on the loans they have taken.

Bivariate Analysis

a. **Term vs Loan Status:** It is observed that there is a higher proportion of borrowers who defaulted on loans in 60 months term than 36 months. Also, the proportion of borrowers who fully paid their loans is higher in 36 months tenure.

b. **Term vs Loan Amount:** It is observed that loan amount is not a deciding factor for default or non-default for both 36 and 60 months tenures.

c. **Term vs Interest Rate:** It is observed that for higher interest rates, the default rate is higher for both 36 and 60 months tenure.

d. **Term vs DTI:** It is observed that comparatively charge-offs are higher when compared with fully paid loans for Debit to Income Ratio.

e. **Grade vs Loan Status:** The bar graph shows that the charged-off increases as the grade decreases.

f. **Employment Length vs Loan Status:** It is observed that employment length does not have any significant impact on loan default.

g. **State vs Charge-Off:** It is observed that more number of borrowers defaulted in the states of Ca, NY, FL & TX.

h. **Purpose vs Loan Status:** It is observed that charged-offs are higher for borrowers of type small businesses.

Recommendations

a. Lending Club should refrain from issuing high interest rate loans for 60 months tenure as the latter is more prone to default by borrowers.

b. Borrowers of type small business have defaulted more on their loans and hence Lending Club should refrain from issuing loans to these category of borrowers.

c. Since the borrowers from the states like A, FL, NY & TX have defaulted more, Lending Club should reduce the no. of loans issued to the borrowers from these 4 states.

d. Lending Club should be cautious while issuing loans to borrowers with a lower grade.