Lending Club Case Study

ML C54

Rohit Pothuru

Business objective

We need to look at the possible loan factors causing default to help the company avoid further losses

Based on the business objective it is prudent to take factors/columns that are considered pre-approval of loan

Factors like employment status , annual income , customer credit status are the most effective in helping us analyze delinquency and improve overall profit position of Lending Club

Methodology

- **Data cleaning** : We should be focused on factors causing loan default before approval of the loan so we can remove other columns. For ex:

a) Funded amount and recovered amount doesn't contribute to the final goal

b) Recoveries and collections also shouldn't matter to our analysis as they are post charge-off

c) Payment dates and payment dues do not contribute to the analysis

- **Aligning with the business objective** : Since we cannot analyze much with ‘Current’ active loans and we only have to look at charged off loans and fully paid loans as a business objective we can remove current loans
- **Handling missing data** : We need to remove null values and replace them with the mode of the data as null data would skew data distribution
- **Handling special characters** : Data needs to be in proper standard format to achieve proper visualizations of data. For ex :

a) In employment length field , ‘<’ and ‘+’ can be replaced by the values itself as data dictinary states 10 means greater than equal to 10 and 0 means less than 1 year

Observations from univariate analysis

- The highest percentage of defaulted loans were B grade loans
- The highest percentage of defaulted loans were customers living in rented houses where the number of defaults were > 700
- One more observation is A grade loans were offered more to homeowners than renters where the number of defaults is < 300
- Highest number of defaults were with customers with employment length greater than 10 years
- Percentage of delinquent loans was higher with shorter term loans(3 years) than longer term loans(5 years)
- Debt consolidation was the major reason put down by the customer for loan application and these category of loans had abnormally high in default rates(> 2500) which was more than 50% of the entire loans applied
- There is a linear relation between annual income of the borrower and the loan amount applied for charged off loans
- Highest number of defaulted loans were issued in the year 2011 and the month of December had a high rate of default than all other months. This could be the after-effects of the mortgage crisis of 2008
- Interest rates for charged off loans were higher than fully paid loans (>13%)
- Loans that were not verfified lead to higher defaults (> 2000)
- Public recorded bankruptcies which were not verified lead to defaults as well
- There is a linear relationship between number of open credit lines and annual income as well
- Customers with higher number of open credit lines and higher incomes were also the ones that defaulted more

Bivariate analysis for probability of defaults

a) The number of total credit lines increase as loan amounts taken increase
b) Annual income and loan amounts are also directly proportional
c) Instalments are higher with high loans which is expected
d) Lower grade(G,E) loans have very high interest rates(>20%) even when loan amount is low
e) Higher amounts invested by investors lead to higher interest rates applied on loans
f) Customers with higher annual incomes seem to have higher revolving credit balance in their accounts
g) B grade loans which exhibited high defaults were given to customers with lower revolving credit balances
h) Highest number of loans defaulted were in the range of dti 10-15
i) Higher default probability when the customer is a homeowner with an income range of >60000
j) Higher defaults when the loan amount is higher(~140000) and is taken for a small business purpose
k) Higher defaults are seen where customer dti > 14 and income ranges from 31k – 85k
l) Higher defaults are seen with employee tenures greater than 6 yrs and loans in 25k – 30k range
m) Smaller unverified loans in the range of 0-10k were also leading to higher defaults