Customer Segmentation based on RFM and predicting Defaulters

Shravan Ahuja

Madhav Sharma
Abstract

• Lending Club is a US based peer to peer lending company which enables borrowers to create personal loans between $1000 to $40000 for a term of 3 or 5 years.
• We used the lending club data set that consists of its current and former customers to get an insight about the risks for investors and the reliability of borrowers.
• Our goal is to
  • use RFM segmentation to identify our most valued customers
  • predict likelihood to default
  • finding the best customers to upsell more loans.
• We also looked at a potential strategy to minimize the losses incurred by charging off loans.
Data Acquisition and Cleaning

• The data was acquired from Lending Club’s website.
• [DATA ACCESS: https://www.lendingclub.com/info/download-data.action]
• The Default variable was created by recoding the loan_status variable; the fully paid loans were coded as Default=0 and default and charged off, were coded as Default=1.
• Distributions were analyzed and suitable transformations were applied to variables with high skewness and kurtosis levels.
• For the predictive models, variables directly dependent on the target variables were not considered.
RFM analysis

• We came up with RFM metric based on Reccency- Issue_date, Frequency- no of accounts, Monetary- loan_amnt.

• We found that most of our customers (>1000) are grouped into mainly 4 RFM Intervals which is inconsistent with the grades assigned by the company based on FICO score.

• These RFM intervals were further resorted to get a clearer picture about likelihood to default.
Predictive Models

• A Logistic Regression model gave interest rate, Log(Annual Income), term and purpose as the important variables.
• Decision Tree made the first split on Interest rate followed by purpose.
• Interest Rate >11.03% = Higher chance of default (or vice versa)
• Annual Income for best customers is the highest and worst customers is moderate, however, it takes a big plunge in the fairly recent customers.
Key Insights

• Customers who take loans for Small Business (27.66%) and Education (20.85%) are more likely to default than others (Mean: 15.12%)
• The data reveals that the Loaning process is very uniform, except for the period between dates in 2010; where risky loans were given to limited number of accounts.
• Debt consolidation is the most popular type of loan (46.49% of all loans).
• The mean interest rate across all RFM groups is 12.6% which is less than the mean of default loaners (13.91%)
Key Insights

• Loans of $73,932,250 were granted to defaulters, of which $47,560,386 were recovered leading to a lost money of $26,371,864. Loss: 35.67%

• While Debt consolidation loans of $28,008,725 were funded at a mean 12.6% ratio to borrowers while $30,971,906 was received in payments. Profit: 10.5%

• Debt Consolidation loans should be marketed to borrowers with high interest rate which will be lucrative to customers with high debt-to-income ratio.
Future Scope

• If 40% of the defaulters are picked with proper screening for the debt consolidation (46.49% of loans); an estimated recovery of $11,603,619 reducing the loss to 29.6%.

• Our RFM analysis indicates that customers with more accounts (frequency) are less likely to default.

• These insights can help Lending club determine their best borrowers to market more loans.
Shravan Ahuja
email: shravan.shuja@okstate.edu
LinkedIn: https://www.linkedin.com/in/shravanahuja/

Madhav Sharma
email: madhav.sharma@okstate.edu
LinkedIn: https://www.linkedin.com/in/madhav-sharma-a627709a/