EDA Final Project Proposal

Lending Club Loan Data Analysis

October 29, 2017

Team Members

- Rahul Raghatate - rraghata@iu.edu
- Syam Herle - syampara@iu.edu
- Saheli Saha - sasaha@iu.edu
- Siddharth Thiruvengadam - sidthiru@iu.edu

Project Description

We are trying to analyze the data of the US peer-to-peer lending company- Lending Club to get insights about their strategies they use to lend loan depending on different features like payment plan, loan status, credit score, installment plan etc. Lending CLub was the first peer-to-peer lender to register its offerings as securities with the Securities and Exchange Commission (SEC), and to offer loan trading on a secondary market. It operates an online lending platform that enables borrowers to obtain a loan, and investors to purchase notes backed by payments made on loans. The Company claims that $15.98 billion in loans had been originated through its platform up to 31 December 2015.

Exploration of Key Business Insights

- Prediction of the risk of charged off for any particular customer before lending the money to the borrower. The metrics associated to activity in the Loan Club are not known because the customer is still a prospect borrower.
- Prediction of probability for a borrower to charge off while he/she is "Current", maybe to prevent the charge off from happening or try to minimise damage.
- Reason people are turning to private lenders instead of banks.
- Distribution of loan applicants through out the United States and the reason behind it.

Data Analysis

We will also perform exploratory data analysis to understand the following:-

- What are the components influencing the interest rate of the loan?
- Factors affecting the loan term
- How are the borrowers who default a loan differ from the borrowers who don't?
- What are the factors that contributes towards loan amount?
- How Loan amount varies with Annual Income? and how the rate varies for annual income?
- Also we will be exploring following factors as part of our analysis for this project.
- How Interest rates varies based on below features:
    1. different states
    2. loan term
    3. purpose of loan application
    4. Profession
    5. Status of loan application
    6. employment length
    7. delinquency

8. different group of borrowers.

Data Description

Kaggle provided dataset(Loan Lending Club data) provides complete loan data for all loans issued through the 2007-2015. Orginally the data has data points and attributes. After cleaning the data we took 887351 observations and 26 attributes. Following are the description of the 26 attributes. **loan_amnt**-The listed amount of the loan applied for by the borrower. If at some point in time, the credit department reduces the loan amount, then it will be reflected in this value.
funded_amnt- The total amount committed to that loan at that point in time.
funded_amnt_inv- The total amount committed by investors for that loan at that point in time.
term- The number of payments on the loan. Values are in months and can be either 36 or 60.
int_rate- Interest Rate on the loan.
installment- The monthly payment owed by the borrower if the loan originates.
grade- LC assigned loan grade.
sub_grade- LC assigned loan subgrade.
emp_length- Employment length in years. Possible values are between 0 and 10 where 0 means less than one year and 10 means ten or more years.
home_ownership- The home ownership status provided by the borrower during registration. Our values are: RENT, OWN, MORTGAGE, OTHER.
annual_inc- The self-reported annual income provided by the borrower during registration.
verification_status- Verification of the customer by the source.
issue_d- The month which the loan was funded
loan_status- Current status of the loan.
purpose- A category provided by the borrower for the loan request.
zip_code- The first 3 numbers of the zip code provided by the borrower in the loan application.
addr_state- The state provided by the borrower in the loan application.
state_full_name- Full name of the states in United States.
delinq_2yrs- The number of 30+ days past-due incidences of delinquency in the borrower's credit file for the past 2 years.
inq_last_6mths- The number of inquiries in past 6 months (excluding auto and mortgage inquiries).
open_acc- The number of open credit lines in the borrower's credit file.
pub_rec- Number of derogatory public records.
total_acc- Total number of credit lines currently in the borrower's credit line.
total_rec_late_fee- Late fees received to date.
tot_cur_bal- Total current balance of all accounts.

Outcomes

- US household debt has been increases by 11% in past 10 years mostly due to nominal income and higher cost of living. As debt has been one of the biggest challenge in the American economic and social development, the insight will be definitely helpful in improving the loan lending business strategies which greatly affects socio-economic development.

References

1. Data source and data dictionary –> https://www.kaggle.com/wendykan/lending-club-loan-data/data
2. https://www.nerdwallet.com/blog/average-credit-card-debt-household/
3. Wikipedia and Lending Club's official website