
Lending Club Case Study

Presented by :

Sameeksha Chiguru & Ankur Dhuriya

Problem Statement

- A consumer finance company which specialises 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.
 - If the Applicant is likely to repay the loan and if loan is rejected for this applicants then its loss of Business to the company
 - If the applicant is not likely to repay the loan and if loan is approved for this applicants then this would lead to Financial loss to the company
 - The company wants to know the potential patterns if a person likely to default
-

Methodology

1. Data Cleaning
 2. Univariate Analysis
 3. Bivariate Analysis
-

Data Cleaning

Steps:

1. Fix rows, columns, Null and Invalid Values
2. Standardize, Transform, Deduplicate

Inference:

- There are 111 columns or variables in the loan data frame.
 - Remove NULL or Missing or Invalid Values
 - 54 columns from the loan has all values NULL is removed
 - Removed columns which non-essential for analysis
 - Remove columns have more than 70 % ONLY has 0 , NA or both
-

-
- Standardize, Transform and Deduplicate
 - Transform Month & Year to standard Datetime Format
 - Remove redundant key attribute – member_id
 - Transform '%' from columns and convert to numeric
 - Convert columns which are Non-numeric to Numeric format
 - Standardize all string columns/attributes to common (Upper) Case
 - Filter out rows to new data frames based on Loan Status for analysis
 - After Data Cleansing the filtered data frames have
 - Charged off – 5627 rows and 41 columns (14%)
 - Full Payment – 32950 rows and 41 columns (83%)
 - Current – 1140 rows and 41 columns (3%)
 - The customer & loan attributes that influence the tendency to default the loan leading to 'credit loss' are considered .
 - The loan status not `Current` (Charged Off & Fully-Paid loan status) .
 - For credit loss analysis ,only data associated with charged off loan status.
-

Univariate Analysis

- Overall Loan status and charged off Loans
- Variables against the number of charged off loans such as
 - Categorical Variables (Nominal & Ordinal)
 - Term, Verification Status, Grade, Sub Grade and Employee Experience
 - Home Ownership and Resident State
 - Purpose
 - Continuous Variables (Interval)
 - Loan Amount
 - Funded Amount , Funded Amount Invested
 - Interest Rate
 - Instalment, Annual Income, DTI
 - Revolving Balance and Revolving Utility %
 - Issue Date, Pulled Credit Date, Last Credit Payment Date

Refer to Univariate Analysis in Jupyter Notebook for the details

Bivariate Analysis

- It includes the below for the charged off loans
 - Home ownership against Verification Status
 - Segmented Annual Income Group (Bins)
 - For Term
 - Verification Status
 - Segmented Annual Income Group
 - Interest Rate Group (Bins)
 - For Term
 - Employee Experience
- Correlation analysis
 - Loan Amount against Funded Amount for different purpose
 - Funded Amount Invested against Instalment for different purpose
 - Loan Payment against Loan Payment Invested for different purpose

Refer to Bivariate Analysis in Jupyter Notebook for the details

Conclusion

- Five states—CA, FL, NY, TX, and NJ—have been observed to approve loans without validating the information. These five states account for nearly 48% of defaulters. Prioritize confirming before approving the loan in these 5 states. Moreover, consumers in these 3 states have been noted to pay installments late.
 - The largest percentage of defaulters reside in Nebraska. Thus, loan applications from this state must be carefully reviewed.
 - Loan Applicants who are obtaining loan for Small Business are the maximum who are Defaulters which is around 27%. The reasons why these Loan Purpose Customers are unable to repay the loan require further examination.
 - The majority of defaulters are loan applicants who got a medium interest rate (10–18%). For these consumers, the interest rate needs to be lowered.
 - Loan applicants who are unable to advance their careers in terms of annual income in relation to work experience and who do not own a home are at the top of the list of defaulters.
-