Exploratory Data Analysis of Lending Club Case Study

Senthil Kumar

& Komal Sinha

The problem

Company

You work for a consumer finance company which specialises in lending various types of loans to urban customers.

Context

Develop a basic understanding of risk analytics in banking and financial services and understand how data is used to minimise the risk of losing money while lending to customers.

Problem statement

The aim 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

Challenges deep-dive

Challenge 1

- Remove columns with null values
- checking columns with single values and removing them
- Dropping all the columns with Single Valued

Challenge 2

- Conversion of Data type
- Strip off unwanted special character
- Handling of Outliner
- Derived Column

Challenge 3

- Understand the variables
- Analysis relations between the variable

Solution

exploratory data analysis

exploratory data analysis is an approach to analyzing data sets to summarize their main characteristics, often with visual methods

Implementation

Univariate Analysis

Analysis:

1. Loan amount
2. Interest Rate
3. Approved Loan Amount percentage
4. Annual Income

Observations:

- The distribution of loan amount, funded amount and funded amount by investor is fairly comparable. In order to use loan_amnt for our additional analysis
- Eighty percent of borrowers received a 100% loan from investors.
- Majority of them have borrowed their loan between 5000 and 15000.
- The majority of interest rates range from 9% to 14.5%.
- The majority of the borrower's yearly income falls between 40k to 80k.
- From January to December, the number of loans granted increases each month, and during the last quarter

Unordered Categorical Variables

Analysis:

1. Loan Status
2. Purpose of Loan
3. Home Ownership

Observations:

- ❖ According to loan status data, 14.3% of loans are charged off, while 85.7% of loans have been fully paid.
- ❖ Based on purpose data, we may infer that the majority of them have taken out loans to pay off credit card debt and consolidate debt.
- ❖ The debt consolidation loan status for paid off loans is likewise high.
- ❖ According to data on property ownership, the majority of them have either rented or mortgaged their homes when taking out loans.
- ❖ The two house owners have high Charged Off loan statuses.

Ordered Categorical Variables

Analysis:

Loan paying term

Observations:

- The majority of them have taken out 36-month loans as opposed to 60-month ones.
- The 60-month loan status with 36 months remaining is likewise high in terms of charged-off loans.

Segmented Univariate Analysis

Analysis:

Home Ownership vs Purpose

Annual Income vs loan status

Observations:

- The majority of borrowers take out loans for less than 14,000, and their charge-off status is similarly high.
- The majority of the borrower's yearly income falls between 40,000 and 60,000.
- The majority of borrowers who obtained loans in order to consolidate their debts have the largest percentage of charges against them, and they also pay the maximum amount in rent.

Bivariate Analysis

Analysis:

Term vs Loan amount

Loan status vs Loan amount

Home Ownership vs Loan amount

Purpose vs Loan amount

Employee Length vs Loan amount

Issue Month vs Loan amount

Observations:

- We can state that the tenure, or 60 months, increases with the loan amount in the term variable. While the 36-month median is 8k, its median is just 15k.
- Verified borrowers receive larger loan amounts than non-verified and source-verified borrowers; that is, over a 10,000 loan amount, all borrowers are confirmed. This is the verification status vs. loan amount relationship.
- Charging off loans resulted in a larger loan amount than fully paid, according to the loan amount vs. loan status

Multivariate Analysis

Analysis:

Correlation between Columns

Observations:

- There is a high correlation between loan_amnt, funded_amnt, funded_amnt_inv, and installment.
- Dti and annual_inc have a negative correlation.
- The percentage of a borrower's monthly gross income that is allocated to debt repayment is known as the debt income ratio.
- This implies that debt is high while annual_inc is low and vice versa

Final Analysis

Observations:

- Those applying for loans with a "60-month" term.
- Candidates whose loan status is "Verified" since they have taken out large loans with a 60-month term
- Candidates who are homeowners who apply for a loan with the intention of consolidating their debts as "Rent"
- Applicants whose annual income is minimal i.e. (0-20000).
- borrowers who accept loans between 0 and 14,000.
- Applicants who receive interest at the rate of 15-20%.
- Those who have applied for a small company loan and whose loan amount exceeds 14,000
- Applicants who has lesser Grade i.e. F<G.
- Candidates with the subgrades F5, G3, and G5.
- Those who are applying for a loan for "home improvement" who make between 60,000 and 70,000

Impact

Mitigate the risk involved while giving loans to the defaulters.