

Lending Club Case Study

Problem Statement

Finding driving variables of defaulter customers(charged off)

Driving Variables Spotted

- Emp Length
- Term
- Interest Rate
- Grade
- Home Ownership
- Annual Income
- Verification Status
- Purpose

Analysis Approach

- Finding driving variables through excel analysis
- Importing required libraries and csv file
- The original data was read through csv file and was reduced to useful data by reducing columns like removing all the unwanted columns and columns of which data would not be available at the time of loan approval.
- Data handling and cleansing variables for null values, missing values and data type conversion
- Data types were changed for analysis for which ever columns were required
- Removed the currently paying records
- Renamed the columns as per the requirements

Analysis Approach

- Starting with describing all the data present in the data
- Checking for outliers in the data by using boxplot
- Removing outliers for 1st std and 2nd std values
- Plotting field values with no outliers
- For univariate analysis we used mean, median, mode and standard deviation formulas and graphs
- For Bivariate analysis, dependency of variables was understood by using correlation and pivot tables.
- Suitable graphs were used to explain the relationships between the variables
- On the basis of our code analysis, observations, insights were drawn and recommendations were suggested

Assumptions

- Emp_length if na then 10.5
- Emp_length if <1 then 0.5
- Emp_length is 1 year the 1 and so on
- Getting 2 derived columns for better bifurcation of data from issue_d: issue_year and issue_month
- Emp_length having value NaN to be changed as mean value of the column emp_length

Data Handling and Cleansing

- Sanity checks for Null Values
- Sanity checks for Missing Values
- Replacing values with suitable values
- Sanity checks for right data types
- Outlier checks through boxplot
- Removing months from term
- Removing % from int_rate
- Renaming term to term_in_months and int_rate to int_rate_in_%
- Removing current loan status from analysis
- Removing outliers with 1st standard deviation as and 2nd standard deviation with respect to rate of interest

Observations

- Most people who take loans live in a mortgage home or rental house
- Most of the applicants have 10+ years of experience
- Most of the charged off loan applicants have taken the loan as a debt consolidation
- People who have taken a loan for a small business are more likely to have charged off
- People with grade B, C and D have the highest charged off count
- People in a mortgaged house, rental or other are more likely to have charged off
- Higher the rate of interest the more likely for the loan to be charged off
- Maximum people have taken loan for debt consolidation
- A5 and A4 have the highest count while B5, B3 and B4 with close values

Insights

- 10,000 is the highest requested loan amount, 5000 is the 2nd highest requested loan amount
- 10,000 is the highest funded loan amount, 5000 is the 2nd highest funded loan amount
- 10,000 is the highest funded loan amount by investor, 6000(approx) is the 2nd highest funded loan amount by investor
- 7.5% is the highest interest rate and 11.5% is the 2nd highest interest rate
- Most of the applicants fall between 25,000 - 60,000 annual income
- People with 36 months term have charged off more. But when you look at the percentage, the percentage of people who have taken the 60 months term loan are more likely to have charged off
- People with average annual income of 60000 rupees have lesser charged off ratio.

Insights

- Over 21212 people have paid the loan while 3188 people have charged off
- Over 11739 people are not verified, 6733 people are source verified while 5928 are verified
- Over 19346 people take loans for 36 months, while 5054 people take loans for 60 months
- 7399 people have grade A and 7511 people have grade B have the most people. 4852 people have grade C with the 3rd highest value
- Payers with average loan amount for emp_length 10+ have better volumes of loans and lesser charged off ratio. Also number of loans has sudden drop from 0. Sudden spike at 10+
- Payers with average loan amount of rounded off figures like 10000, 12000 and 6000 have lesser charged off ratio than approximated figures

Recommendations

- People with grade A can be provided with loans easily as they have significant numbers of loan payers and less defaulters.
- Among the home ownership types, we can provide loan easily to mortgage category than to rent and own category
- Among the purpose variable, we can provide easy loan to debt consolidation category
- The people with annual income of average 60000 rupees should be granted loan as they lead to less charged off percentage when compared to other category
- If any party asking for 10000 rupees loan amount then it should be funded easily as they are unlikely to charged off.
- Payers with employee_length 10+ should be preferred