CS 59 Lending Club Case Study

Veda Vamsi
Vaibhav Kumar

Approach Taken

Objectives

Data Cleaning

Univariate Analysis

Multivariate Analysis

Additional Details

Objectives

- If one is able to identify risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. Identification of such applicants using EDA is the aim of this case study.

- The company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default.

Data Cleaning

- Remove columns which are null

- Remove Identifier columns and single valued columns

- Remove investor related columns

- Now check null values in remaining columns as percentage of all values

- Fill the missing values with mode or remove the rows which still have few columns as null values, if percentage is very small

- Remove current loans data and focus primarily on 2 columns i.e. loan amount and annual income. Then remove outliers for the 2 cols

Univariate Analysis

- Use charged off loans data to observe defaulting reasons

- Home ownership, purpose of loans taken can help to see possible defaults reasons

- Create 2 bins for annual income and interest rate and observe charged off loans data against the 2 bins

- Similarly check for emp length also against charged off loans, and we find :

- People living on rent have higher default rates

- People using loans to reduce other obligations belong to the highest category from Purpose perspective

- People with employment length of 10 years have higher defaulting rates

- People getting loans at 13-17 % default more

- People having annual income between 30000 USD and 60000 USD default more

Multivariate Analysis

- 2 columns prioritized – annual income and loan amount and loan_status to be used as a categorical variable

- Using annual_income, home ownership, purpose, loan amount and interest rate we can find :

- People defaulting more when –
    - having annual income as 60000 - 70000 USD and home is mortgaged
    - having higher loan amounts as compared to others in the same income group
    - owning home as "MORTGAGE' and have a loan amount of over 14000 USD default more
    - taking a loan for small business and amount is over 14000 USD
    - having an interest rate of 15 - 17.5 % and loan amount of 30000 - 35000 USD

Additional Details

- Github link :
https://github.com/vaibhavk14/lendingcase