Lending club case study

Priyesh Raj

Raj.Priyesh@gmail.com

Business problem

- Build a risk profile of individual
- Identify what are the strong drivers for loan default.

Data cleaning and standards

- Data loaded in a data frame and cleaned.
- Null columns dropped.
- Null values in rows filled with appropriate values.
- Columns which had no strong role in decision making were dropped from the data frame.
- Outliers identified and removed using 95 Quantile values.
- Strings were modified to integers for proper usage.

Univariate Analysis

- Columns identified for univariate analysis.
  - Home ownership.
  - Purpose of loan
  - Term of loan
  - Grade of employee
  - Public record of defaults or cases.
  - Annual income group
  - Interest rate group

Results from analysis

- ***It seems that most of the default occurs when***
  - When individual is in rented home
  - Debt is taken for debt reconciliation
  - The term is 36 Month
  - Income rate falls in 13-17% group
  - Income falls in bracket of 30-60K
  - There is 1 or more case running on the individual.

BiVariate Analysis

- Columns identified for bivariate analysis.
  - Income with house ownership status
  - Income with purpose of loan
  - Loan amount with purpose of loan
  - Loan amount with house ownership

Results from analysis

- **So, combining two variables, few interesting insight came out**
  - Loan borrowing is higher for people living in rented home or mortgaged home.
  - Loan with Small business and debt consolidation have higher default rate.
  - Individual with higher income group generally takes loan for home improvement.

Risk profile based on analysis

- Individual with
  - income bracket 30-60K with
  - rented home and
  - loan purpose of debt consolidation or small business,
  - if offered a 13% or higher interest rate
- will have higher chances of default on their loan.