LENDING CLUB CASE STUDY

BY SANTOSH KUMAR
SANTOSH.ASSISI@GMAIL.COM

PROBLEM STATEMENT

- The consumer finance company facilitating personal loans, business loans, and financing of medical procedures. Lending loans to ‘risky’ applicants is the largest source of financial loss (called credit loss). Credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed. In other words, borrowers who default cause the largest amount of loss to the lenders.

- The company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.

- We need to identify these 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.

ANALYSIS APPROACH

- The first step after loading the data should be to clean the data, identify the proper datatypes. Remove records with null values.
- Analyze the dataset and identify the key columns/driving columns which will have impact in the case study. In our example, we need to find the columns and also derive the columns which will directly or indirectly tell about the defaulting probability.
- With the help of several graphs, identify the relationship and try to get the insight.
- Draw the univariate, bivariate analysis graph . This gives more details insight in easy way.

RESULT:- UNIVARIATE ANALYSIS

- Created below graph:-

  Average distribution of loans across category and experience level

  Distribution of loans across category and dti

  Distribution of loans across category and loan to annual income

  Distribution of loans across category and Grades

Analysis :-

people with less experience have high chance of default.

People lying in medium dti range have high chances of default

People who have high loan to annual income ratio are at high risk of defaulting.

Grades B,Cand D are at high probablity of defaulting.

RESULT:- BIVARIATE ANALYSIS

- Created below graph:-

    Distribution of Verified and Not verified Source Income on DTI

    Distribution of Verified and Not verified Source Income on Loan to Annual Income

    Distribution of Verified and not verified Source Income on Experience Level

    Distribution of Verified and not verified Source Income on Grades

    Distribution of Rented, Mortgaged, Owned asset ownership on DTI

    Distribution of Rented, Mortgaged, Owned asset ownership on loan to annual income ratio

    Distribution of Rented, Mortgaged, Owned asset ownership on experience level.

    Distribution of Rented, Mortgaged, Owned asset ownership on grades

ANALYSIS:-

- average default rate across all categories is 14.4%.
- The number of loan applicants is increasing every year, 2011 being the highest.
- People with 1 year and 10 year experience are taking loans more frequently.
- Top 6 loan category:- debt consolidation credit card other home improvement major purchase small business
- People with less experience have high chance of default.
- People lying in medium dti range have high chances of default
- People who have high loan to annual income ratio are at high risk of defaulting.
- Grades B, C and D are at high probability of defaulting.
- For verified source of income. High loan to income ratio has highest probability of defaulting.
- For verified and not verified source of income. The Junior experience level are more towards defaulting.
- For rented and house owner. More the loan to income ratio, more the chances of defaulting.