Exploratory Data Analysis on Loan Data

Atanu Basak

Here is the analysis on loan data, which give the idea on how other attributes effect loan defaulters. I have taken the different attributes like loan purpose, revolving balance, fico, installment and interest rate into consideration.

At the regression analysis it shows among the loan purpose credit card, debt consolidation and small business loans have significant effect on the defaulter data. Among the other attributes fico and interest rate also have significant effect on the loan data for defaulters.

**What’s missed during the analysis?**

There are some restrictions on the PII data as loan data is very sensitive, that give the behavior of the individual. There could be further analysis need to be done to identify the defaulter. Payment Analytics and consumer bureau attributes are missed in the analysis. If those are taken into consideration we could have better picture to identify the actual reasons for loan defaulters.

**Were there any assumptions made felt were incorrect?**

Interest rate and fico are highly correlated attributes, but I have included both of them in the regression analysis, here I took the assumption that they are not related to each other, otherwise I should have only selected one attributes among them to see the effect on the defaulter data.

**What are challenges faced during the analysis?**

One the of attribute, the purpose of the loan is categorical and have 7 categories on it, so I have to do the one hot encoding to create separate fields for each attribute with flagging. This increased the complexity of the regression analysis.