Project 1. Report

We have started our analysis of loan data by reading data in R

We found that there are few variable which is to be changed and

Cleaned out for analysis purpose. We have omitted all NA values

From the data and cleaned the data also we have changed Fico.Score

variable to Median Fico.Score for our analysis purpose.

we found that there are some irrelevant observations in the variable

Amount funded by the investors we set them as NA and after that we have

Removed those NA values.

We have gone through different plots like box plots , scatter plots

to check the trends of other variables with our target variable

Interest. Rate, we are also plotted the correlations .

We got the following observations

Interest. Rate is normally distributed

higher loan length has higher interest rate

Median interest rate is almost equal for each factor with little difference of Employment Length variable.

we have got skewed graph Monthly. Income and Revolving.Credit.Balance we transform thses variables to log and found that in scatter plot Monthly.Income has increasing trend and Revolving.Credit.Balance has decreasing trend with Interest.Rate.

Also form the box plot of Fico.Score and Interest.Rate it is clear that higher Fico.Range has lower Interest.Rate this is one of the important obseration along with the Loan.Length for higher time has higher Intrest.Rate .

Then we set our data for regression

We have divided the loan data into training as well as test set.

then we run a linear regression on training set as well as test set

on training set we found that few variables are not significant

so taking graphichs and significance and correlations and

multicollinearity into account we dropout some variable and run new

regression model for both the set , we found that for the new model

all the variables are significant and there are no mullticollinearity

problems also the coefficients for both train and testset are equal and

R^2 = 0.7476   for train set and R^2= 0.7673 for test set

and p value < 2.2e-16 shows the goodness of fit of the model.

Conclusion:-

Our analysis suggests that there is a significant,

negative association between interest rate and FICO score.

Our analysis estimates the relationship using a linear model

Relating FICO score with interest rate.

We also observed that other variables such as loan length and Amount funded and Inquiries.in.the.Last.6.Months are associated with both interest rate and FICO score.

Including these variables in the regression model, relating interest rate to FICO score improves the model fit, but does not remove the significant negative relationship between interest rate and FICO score.

High FICO SCORE indicate the low interest rate