{ ‘Name’ : ‘ Subodh D ’ , ’Class’ : ‘ MSc CA Data Science ’ ,‘PRN’ : 18030142043 ,
‘Email’ : ( ‘sus1842043@sicsr.ac.in , replytosubodh@gmail.com' )}

## Loan Approval Prediction -

## Approach -

2 files were given - one was of approved loans and another was of rejected loans.
I have deleted some blank records in excel which were causing problem for frame loading.
I used excel to understand the data and deviations of data.
Cleared unwanted text using programming and manually.
I visualised overall data and I have chosen only few columns from approved loans data as rejection data has few dimensions.
I chose training and testing data based on date variables constraints stated.
Approval of loan depends on different factors basically your income and credit history.It has multiple other factors which include your co-borrower, previous loans, job profile and experience.
Some people take loan to repay existing loan which has higher interested.Some loan comes with different offer from small investor banks.Some people get pre approved offers on loans

## Data Wrangling and Feature selection -

Common features are debt to income ratio, state living, employment length and of course amount.They are the features common in the both data frames.
State names and employment length were in the form of the object so I have converted into float and int data types.
Risk score could have been useful feature but similar feature data was not available in approved loan data column.FICO -high & low and Vantage score columns had only null values.
Rejected loans had record of single borrower so investigated the possibility of any other type exist in the approved loan data.
I have mapped unique values present in both data frame relevant columns with same variables across similar data frame columns.

## Models and Accuracy-

I have tried different models to get accuracy with and with out 10 cv method.
The distribution is very skewed in almost all columns.
There isn’t much correlation between the features itself but collectively they impact except the states mentioned.
The states columns could have been really useful if other information is available.
Example - in states like California living expenses are more compared to other middle situated states in USA and location could also indicate the risk of loan that’s why most loans are approved from California and most rejected loans are also from California.