SCOPE OF PROJECT: Using machine learning techniques we can predict the likelihood of loan status in given test data.

1. Loading all required library

2. Data import Initial setwd command will set current working directory. The file in the format of csv(comma separated version) loaded to R in data.frame format. str command will give observation,data type, name, few sample of variables.

3.Data exploration

4.Finding missing value “NA” using “is.na()”.

A histogram is a graphical representation of the distribution of numerical data. It is an estimate of the probability distribution of a continuous variable.To contruct histogram,the first step is to “bin” the range of values—that is, divide the entire range of values into a series of intervals—and then count how many values fall into each interval.

dpylr package is for efficient data manipulation.

5.Random forest is ensemble learning method for classification,regression. It will create n different tree of subset of train and predict.