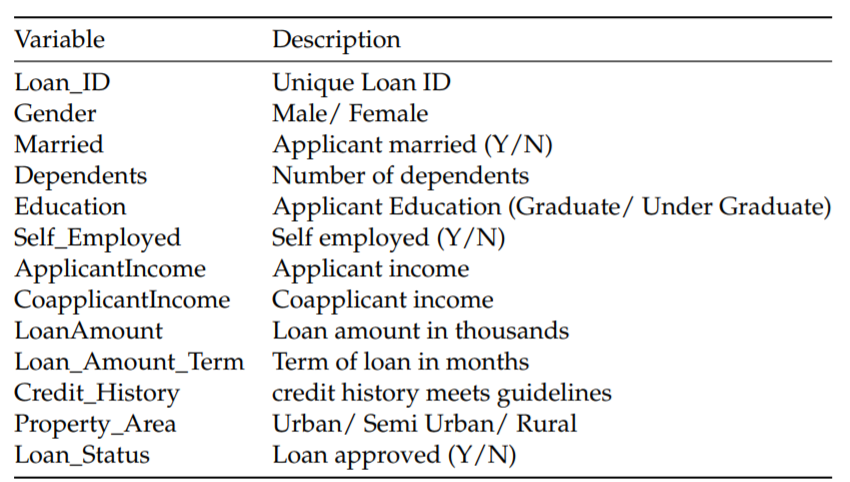
A Company wants to automate the loan eligibility process (real time) based on customer detail provided while filling online application form. These details are Gender, Marital Status, Education, Number of Dependents, Income, Loan Amount, Credit History and others. To automate this process, they have given a problem to identify the customers segments, those are eligible for loan amount so that they can specifically target these customers. Here they have provided a data set.



1) **Describe** the training data set (File name: train.csv) and draw the **Box Plot** for variable Applicant Income by variable Education of training data set.

2) **Compute** Total Income by adding both ApplicantIncome and CoapplicantIncome and plot the Histogram (bins of 20) to understand the extreme values.

3) **Construct a frequency table** for Credit History Vs. Loan Status

3) **Find the missing** values for the given data set

4) **Imputing the missing values** with mean for continuous variable and Imputing Missing mode for categorical variables.

5) Is there any need to apply **stand scalar**? If yes apply and see the result.