#2 Loan Prediction

Load the necessory packages

Load the data

Descriptive analysis - numerical values

->Data.describe()

->Found that some column has missing values . count of first coulmn doesnt match with other coulmns which means there is some missing values

Frequency distribution - categorical values

-> data.column\_name.value\_counts()

It produce count of each values in that column

->data.Property\_Area.value\_counts()

->data.Loan\_Status.value\_counts()

Found that the data for each status is not equal

Use histogram to understand the data

->data.ApplicantIncome.hist(bins= 50)

Most of the applicant income lessthan 10000

-> data.boxplot('ApplicantIncome' , by='Education')

Found that lot of outliers in applicant income for graduate compare to non graduate.

->temp2 = data.pivot\_table(values='Loan\_Status' , index=['Credit\_History'],aggfunc=lambda x:x.map({'Y':1,'N':0}).mean())

To understand the probability of loan status yes when the person have credit history .and vice versa

Data Munging

This is the step we were find the solution for the problems we found

->Fixing thr missing values

data.LoanAmount.fillna(data.LoanAmount.mean() , inplace=True)

Fill with mean()

data.Self\_Employed.fillna('No' , inplace=True)

Fill with ‘No’ becoz 86% data are in No.