# Data Set:

ABC company is a NBFC & deals in all kinds of home loans. They have presence across all urban, semi urban and rural areas. Customer first applies for home loan and after that company validates the customer eligibility for loan.

ABC 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 provided a dataset to identify the customers segments that are eligible for loan amount so that they can specifically target these customers.

# Data Dictionary

**Train file:**CSVcontaining the customers for whom loan eligibility is known as 'Loan\_Status'

**Variable**

**Description**

1. Loan\_ID: Unique Loan ID
2. Gender: Male/ Female
3. Married: Applicant married (Y/N)
4. Dependents: Number of dependents
5. Education: Applicant Education (Graduate/ Under Graduate)
6. Self\_Employed: Self employed (Y/N)
7. ApplicantIncome
8. CoapplicantIncome
9. LoanAmount: Loan amount in thousands
10. Loan\_Amount\_Term: Term of loan in months
11. Credit\_History: credit history meets guidelines
12. Property\_Area: Urban/ Semi Urban/ Rural
13. Loan\_Status: (Target) Loan approved (Y/N)

**Validation file:** CSVcontaining the customer information for whom loan eligibility is to be predicted

**Variable**

**Description**

1. Loan\_ID: Unique Loan ID
2. Gender: Male/ Female
3. Married: Applicant married (Y/N)
4. Dependents: Number of dependents
5. Education: Applicant Education (Graduate/ Under Graduate)
6. Self\_Employed: Self employed (Y/N)
7. ApplicantIncome
8. Coapplicant income
9. LoanAmount: Loan amount in thousands
10. Loan\_Amount\_Term: Term of loan in months
11. Credit\_History: credit history meets guidelines
12. Property\_Area: Urban/ Semi Urban/ Rural

# Submission file format:

1. Loan\_ID: Unique Loan ID
2. Loan\_Status: (Target) Loan approved (Y/N)

# Task:

Design a model that will be able to predict if customer should be eligible for loan or not.

# Deliverables:

Provide the following:

1. The source code you used to build the model and make predictions. (You are

free to use any language and any open-source package/library)

2. A .csv file containing the predictions of the validation data. You can add the target

column (‘Loan\_Status) to the validation data or simply provide it alone with the Loan\_ID

column.

# Briefly answer the following questions:

1. Describe your model and why did you choose this model over other types of models?

2. Describe any other models you have tried and why do you think this model performs better?

1. How did you handle missing data?
2. How did you handle categorical (string) data?
3. How did you handle unbalanced data?
4. How did you test your model?