BCCL DATA SERVICES

**BCCL Data Services** is a global organization with more than 10,000 clients and 20,000 associates worldwide and takes pride in its mission to enable clients to achieve best-in-class results. The company is highly regarded for its financial services technology and services innovation, including award-winning solutions for mobile and online banking, payments, risk management, data analytics and core account processing.

BCCL Data Services is helping its clients to push the boundaries of what's possible in financial services, delivering deep expertise and innovative solutions to help financial institutions, businesses and consumers move and manage money faster and with greater ease than ever before.

XYZ Bank deals in all home loans. They have presence across all urban, semi urban and rural areas. Customer first apply for home loan after that company validates the customer eligibility for loan.

XYZ bank wants to understand the usability of the customer details it has collected using online / offline application form. Hence, they need to automate the loan eligibility process (real time) based on customer detail provided to them. XYZ Bank signs a project with BCCL Data Services in order to make their customer details more informative and identify the customer segments, those are eligible for loan amount so that they can specifically target these customers. XYZ Bank will share all the customer details in a comma separated file (csv) with BCCL Data Services which will then be analysed.

You are a senior data analyst from BCCL Data Services and are assigned the task to analyse the submitted data by XYZ Bank. **Write an algorithm using the train and test data set and provide a data that should generate a prediction csv file.**

**DATA FILES (3 files)-**

| **File Name** | **Description** | **Format** | **Size** |
| --- | --- | --- | --- |
| test\_data.csv | the test data set | .csv | 25.8 (kb) |
| train\_data.csv | the training data set | .csv | 5.1 (kb) |
| train\_prediction.csv | a prediction data set consist of loan status for train data set | .csv | 0.49(kb) |

test\_data.csv and test\_data.csv file contain the following data of customer

1. loan\_id - Unique Loan ID
2. gender - Male/ Female
3. marital\_status - Applicant married (Yes/No)
4. dependents - Number of dependents
5. qualification - Applicant Education (Graduate/ Under Graduate)
6. is\_self\_employed - Self employed (Yes/No)
7. applicant\_income - Applicant income
8. co\_applicant\_income - Co-applicant income
9. loan\_amount - 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

Your prediction csv file should contain the following information:

1. loan\_id - Unique Loan ID
2. status - Loan approved (Y/N)

**EVALUATION CRITERIA**

Score is calculated with the following formula:

**Score = Number of correct predictions / Total records\*100**

Note: *Submission implementation will be evaluated based on accuracy.*

**GUIDELINES**

1. You must submit your train\_prediction.csv in the format given in problem statement. In addition to your output, you will also have to submit your algorithm as well.
2. You can use any tools or libraries for building your solution. There is no restriction on the tools that you can use.