

# EDA CASE STUDY:

Problem Statement:

**To identify patterns** which indicate if a client has difficulty paying their instalments which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate

# ABOUT DATASET

- Two types of risks are associated with the bank's decision:
- If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
- If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company.
- About Data:
- Application Data: contains all the information of the client at the time of application.  
The data is about whether a **client has payment difficulties**.
- Previous Application Data: contains information about the client's previous loan data. It contains the data whether the previous application had been **Approved, Cancelled, Refused or Unused offer**.
- Column Data: which give brief about each column in above 2 dataset.

# Outcomes of Observation:

- 1. People having Lower Secondary education are very less and it also has maximum % of payment difficulties- around 12%. Hence, client with education type as 'Lower Secondary' are the driving factors for Loan Defaulters.
- 2. In application data study we see Female were most who were likely to apply and get approval of loan but also it has counter parts that male are more likely to be defaulter that is around 9.5% .
- 3. People having CREDIT\_INCOME\_RATIO > 30% are driving factors for loan defaulter
- 4. nuclear family(2-3 people in family) get highest approval.
- 5. Get as much as clients from housing type 'With parents' as they are having least number of unsuccessful payments.