

# **Assignment 1 :-**

## **High Risk vs Low Risk Applicants.**

**A credit score is based on an individual borrower's credit history, that is, the number of credit accounts, total debt, repayment history and enquiries he/she has made while seeking a loan. Using the credit score, a lender evaluates the probability of the borrower repaying debts/loans on time.**

**While a high score increases your chances of getting credit (like a loan or credit card) and that too at attractive interest rates, a lower score can make lenders wary of lending to you.**

**So, based on the given past record data of a bank granting its customers the loan in terms of their credibility, where every individual is categorised as either a 'High Risk Applicant' or a 'Low Risk applicant'.**

**( 'High Risk Applicant'=1 & 'Low Risk applicant'=0).**

## **My Analysis Report :-**

**So, based on my analysis over the given data, I came across these various insights:-**

- **More Males apply for loan than Females**
- **People with a maximum of 4 years of employment gets loan mostly.**
- **People with at least 1 years of employment gets loan mostly.**
- **Mostly Single people apply for loans compared to others.**
- **Banks mostly prefer Skilled and employed people for granting loans.**
- **Mostly people take loans to buy Electronic Equipment.**

## **These are some characteristics of a High Risk Applicants:**

- **Younger applicants (<35 yrs old) are more likely to become a high risk applicant, compared to the Older applicants (>35 yrs old).**
- **People who take loans for 24 months (atleast), generally fall under the category of a high risk applicant.**
- **Single people are most likely to become a high risk applicant.**
- **People who have at least one existing loan in the bank are most likely to become a highly risk applicant.**

## Observation by using Machine Learning Algorithms in the given Data

- The given Data is around 73% accurate. Which means it can classify the applicants on the basis of their credibility at ~73% accuracy.
- There are 14 Numerical columns and 11 Categorical columns in the given data.
- The Principle amount column is left screwed and has around 40% of outliers present in it.

## Conclusion

**The above analysis might be crucial for the bank,  
Although the given Data is ~73% accurate which might not  
give accurate predictions every time.**