**Problem Statement – I**

**Introduction**: This case study aims to give you an idea of applying EDA in a real business scenario. In this case study, apart from applying the techniques that we have learnt in the EDA module, we will also develop a basic understanding of risk analytics in banking and financial services and understand how data is used to minimise the risk of losing money while lending to customers.

**Business Understanding**: The loan providing companies find it hard to give loans to the people due to their insufficient or non-existent credit history. Because of that, some customers use it as their advantage by becoming a defaulter. Suppose you work for a consumer finance company which specialises in lending various types of loans to urban customer. You have to use EDA to analyse the patterns in the data. This will ensure that the applicants capable of reapplying the loan are not rejected.

When the company receives a loan application, the company has to decide for loan approval based on the applicant’s profile. Two types of risk are associated with the bank’s decision:

* If the applicant is likely to repay the loan, then not approving the loan 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.

The data given contains the information about the loan application at the time of applying for the loan. It contains two types of scenarios:

* **The client with payment difficulties**: He/she had late payment more than X days on at least one of the first Y instalments of the loan in our sample,
* **All other cases**: All other cases when the payment is paid on time.

When a client applies for a loan, there are four types of decisions that could be taken by the client/company:

1. **Approved**: The company has approved loan Application.
2. **Cancelled**: The client cancelled the application sometime during approval. Either the client changed her/his mind about the loan or in some cases due to a higher risk of the client he received worse pricing which he did not want.
3. **Refused**: The company had rejected the loan(because the client does not meet their requirements etc.).
4. **Unused offer**: Loan has been cancelled by the client but on different stages of the process.

**Business Objectives**:

This case study aims 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, etc. This will ensure that the consumers capable of repaying the loan are not rejected. Identification of such applicant’s using EDA is the aim of this case study.

In other words, the company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are string indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.

To develop your understanding of the domain, you are advised to independently research a little about risk analytics – understanding the types of variables and their significance should be enough)