Problem Statement

Introduction
Solving this case study will give you an idea of how real-life business problems are solved using EDA. In this case study, you will apply the techniques that you have learnt in EDA. You will also develop a basic understanding of risk analytics in banking and financial services and understand how data minimizes the risk of losing money while lending it to customers.

Business Understanding

You work for a consumer finance company that specialises in providing various types of loans to urban customers. When the company receives a loan application, it has to decide whether to approve or reject it based on the applicant's profile. 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., they are likely to default, then approving the loan may lead to financial loss for the company.

The data given below contains information about past loan applicants and whether they 'defaulted'. The aim is to identify patterns indicating that a person is likely to default, which may be used to deny the loan, reduce the loan amount, lend (to risky applicants) at a higher interest rate, etc.

In this case study, you will use EDA to understand how consumer attributes and loan attributes influence the tendency of defaulting.

The following image depicts the decisions that could be undertaken by the firm.
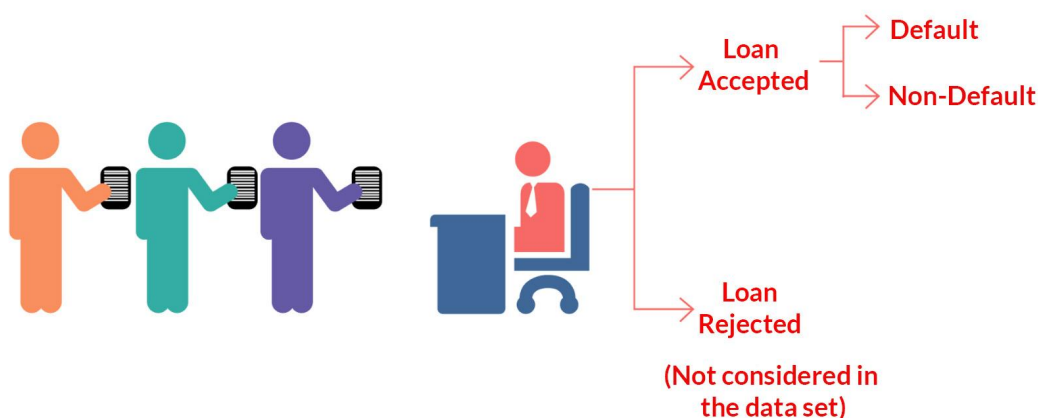
When a person applies for a loan, there are two types of decisions that could be taken by the company:

Loan accepted: If the company approves the loan, there are three possible scenarios, as described below:

Fully paid: The applicant has fully paid the loan (the principal and the interest amount).

Current: The applicant is in the process of paying the instalments, i.e., the tenure of the loan is not yet completed. These candidates are not labelled as 'defaulted'.

Charged-off: The applicant has not paid the instalments in due time for a long period, i.e., they have defaulted on the loan.

Loan rejected: The company had rejected the loan (because the candidate does not meet their requirements, etc.). Since the loan was rejected, there is no transactional history of those applicants with the company; so, this data is not available with the company (and thus, in this data set).

Business Objectives

This company is the largest online loan marketplace facilitating personal loans, business loans, and the financing of medical procedures. Borrowers can easily access lower interest rate loans through a fast online interface.
Like most other lending companies, lending to 'risky' applicants is the largest source of financial loss (called credit loss). Credit loss is the amount of money lost by the lender when the borrower refuses to pay or absconds with the money owed. In other words, borrowers who default cause the biggest losses to lenders. In this case, customers labelled as 'charged-off' are the 'defaulters'.

If one is able to identify these risky loan applicants, then such loans can be reduced, thereby cutting down the amount of credit loss. Identification of such applicants 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 strong 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).

Results Expected

The following are the anticipated results from the learners' end:

Write all your code in one well-commented Python file; briefly mention the insights and observations from the analysis.
Present the overall approach of the analysis in a presentation:
Mention the problem statement and the analysis approach briefly.
Explain the results of univariate analysis, bivariate analysis, etc., in business terms.
Include visualisations and summarise the most important results in the presentation.