Problem Statement:-

You work for a **consumer finance company** which specialises in lending various types of loans to urban customers. When the company receives a loan application, the company has to make a decision for loan approval 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. he/she is likely to default, then approving the loan may lead to a **financial loss** for the company

The data loan.csv.zip contains information about past loan applicants and whether they 'defaulted' or not. The aim is to identify patterns which indicate if a person is likely to default, 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.

Business Objectives

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

Results Expected

1. Write well-commented Python code; briefly mention the insights and observations from the analysis
2. Insert the data in MongoDb for future usage.
3. Present the overall approach of the analysis in a presentation:
    - Mention the problem statement and the analysis approach briefly
    - Include visualisations and summarise the most important results.