## **Loan Default Prediction Dataset**

**Project Goal:**

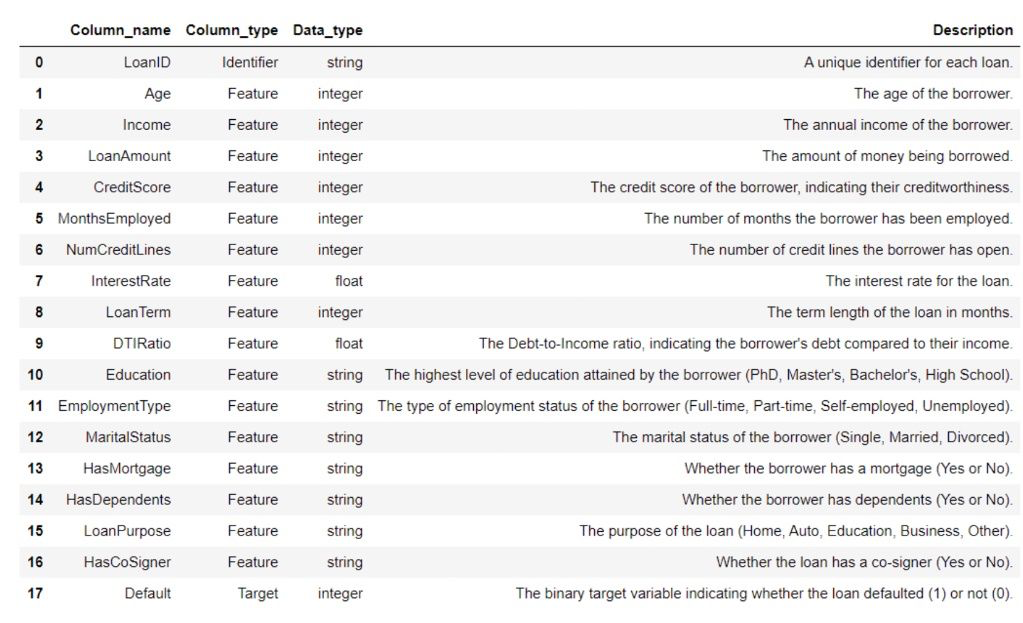
To develop a machine learning model that predicts the risk of a borrower defaulting on a loan and recommends personalized recovery actions.

**About Dataset:**

The **Loan Default Prediction Dataset**focuses on predicting whether a borrower will default on a loan based on various demographic, financial, and loan-related attributes.

The dataset consists of **255,347 records** and **18 columns**, including both numerical and categorical variables. Each record represents a loan applicant with details related to their personal profile, employment, financial status, and loan information. The target variable, **“Default”**, indicates whether the loan was defaulted (1) or not (0).

**Overview:**



**Deep-dive into financial parameters:**

**1. Financial Health Indicators:** These features directly measure a borrower's financial standing and ability to handle debt.

* **Income:** Higher income generally means a greater capacity to repay loans, lowering the risk of default.
* **CreditScore:** This is a highly significant predictor. A high credit score indicates a history of responsible borrowing and timely repayments, suggesting a lower default risk. Conversely, a low score signals higher risk.
* **DTIRatio (Debt-to-Income Ratio)**: This crucial metric shows how much of a borrower's monthly income goes towards paying off debts. It's calculated as:  
  DTI = (Gross Monthly Income) / (Total Monthly Debt Payments​)  
  A lower DTI is better, as it suggests the borrower isn't over-leveraged and has enough income left over to handle new loan payments. A high DTI is a major red flag.
* **LoanAmount:** A very large loan amount relative to the borrower's income can increase the financial strain, potentially raising the default risk.

**2. Stability and Responsibility Indicators:** These features provide context about the borrower's life stability and sense of responsibility, which can indirectly signal their reliability.

* **MonthsEmployed & EmploymentType:** A long and stable employment history (e.g., many months in a full-time job) is a strong positive signal. It suggests a consistent and reliable source of income. Frequent job changes or unemployment are risk factors.
* **Age:** While not always a direct cause, age can be a proxy for financial stability and experience. Older borrowers may have more stable careers and assets.
* **MaritalStatus & HasDependents:** These factors can influence financial stability. For example, having dependents might increase financial strain, but a married applicant might have a dual-income household, which could lower risk.
* **HasMortgage:** Having a mortgage and paying it on time is a strong indicator of financial responsibility and creditworthiness.
* **Education**: Higher levels of education often correlate with higher, more stable incomes, which can lower the risk of default.

**3. Loan Characteristics:**

* **InterestRate:** Higher interest rates lead to higher monthly payments, which can strain a borrower's budget and increase the chance of default. Rates are often higher for riskier borrowers in the first place.
* **LoanTerm:** A longer loan term means lower monthly payments but more interest paid over time. A very long term might be taken by someone who can't afford higher payments, indicating potential risk.
* **LoanPurpose:** The reason for the loan can be indicative of risk. For example, a loan for a business venture might be riskier than a loan for a home improvement project.
* **HasCoSigner:** The presence of a co-signer (someone who agrees to pay the debt if the primary borrower cannot) significantly reduces the risk for the lender, as there is a second person legally responsible for the repayment.
* **NumCreditLines:** This shows how much credit the borrower already has access to. A very high number of open credit lines could indicate that the borrower is over-extended and relies heavily on debt.