<https://www.kaggle.com/datasets/lorenzozoppelletto/financial-risk-for-loan-approval>

**About Dataset**

**Synthetic Dataset for Risk Assessment and Loan Approval Modelling**

This synthetic dataset comprises 20,000 records of personal and financial data, designed to facilitate the development of predictive models for risk assessment. It serves two primary purposes:

1. **Risk Score Regression:** To predict a continuous risk score associated with each individual's likelihood of loan default or financial instability.
2. **Binary Classification:** To determine the binary outcome of loan approval, indicating whether an applicant is likely to be approved or denied for a loan.

The dataset includes diverse features such as demographic information, credit history, employment status, income levels, existing debt, and other relevant financial metrics, providing a comprehensive foundation for sophisticated data-driven analysis and decision-making.

The dataset includes the following columns:

* **ApplicationDate**: Loan application date
* **Age**: Applicant's age
* **AnnualIncome**: Yearly income
* **CreditScore**: Creditworthiness score
* **EmploymentStatus**: Job situation
* **EducationLevel**: Highest education attained
* **Experience**: Work experience
* **LoanAmount**: Requested loan size
* **LoanDuration**: Loan repayment period
* **MaritalStatus**: Applicant's marital state
* **NumberOfDependents**: Number of dependents
* **HomeOwnershipStatus**: Homeownership type
* **MonthlyDebtPayments**: Monthly debt obligations
* **CreditCardUtilizationRate**: Credit card usage percentage
* **NumberOfOpenCreditLines**: Active credit lines
* **NumberOfCreditInquiries**: Credit checks count
* **DebtToIncomeRatio**: Debt to income proportion
* **BankruptcyHistory**: Bankruptcy records
* **LoanPurpose**: Reason for loan
* **PreviousLoanDefaults**: Prior loan defaults
* **PaymentHistory**: Past payment behavior
* **LengthOfCreditHistory**: Credit history duration
* **SavingsAccountBalance**: Savings account amount
* **CheckingAccountBalance**: Checking account funds
* **TotalAssets**: Total owned assets
* **TotalLiabilities**: Total owed debts
* **MonthlyIncome**: Income per month
* **UtilityBillsPaymentHistory**: Utility payment record
* **JobTenure**: Job duration
* **NetWorth**: Total financial worth
* **BaseInterestRate**: Starting interest rate
* **InterestRate**: Applied interest rate
* **MonthlyLoanPayment**: Monthly loan payment
* **TotalDebtToIncomeRatio**: Total debt against income
* **LoanApproved**: Loan approval status
* **RiskScore**: Risk assessment score

Goals:

1. **Risk Score Regression:** To predict a continuous risk score associated with each individual's likelihood of loan default or financial instability.
2. **Binary Classification:** To determine the binary outcome of loan approval, indicating whether an applicant is likely to be approved or denied for a loan.