**Analyze and Predict Default Risk Based on Financial and Demographic Factors**

* **Goals of the Project**: The main goal is to build a predictive model to estimate the probability of loan default based on financial and demographic factors. By understanding key factors influencing default risk, the model aims to help financial institutions improve credit risk assessment and make data-driven lending decisions.
* **Data Sources Used**: The dataset is sourced from Kaggle, specifically the "Loan Default Prediction Dataset." This dataset contains various financial and demographic variables related to borrowers, which are essential for analyzing and predicting default risk.
* **Data Overview**: The dataset includes details about each loan, such as age, income, credit score, and employment information, which are useful in assessing credit risk.
* **Structure**:
  + LoanID: A unique identifier for each loan
  + Age: The age of the borrower
  + Income: The annual income of the borrower
  + LoanAmount: The amount of money being borrowed
  + CreditScore: The credit score of the borrower, indicating their creditworthiness
  + MonthsEmployed: The number of months the borrower has been employed
  + NumCreditLines: The number of credit lines the borrower has open
  + InterestRate: The interest rate for the loan
  + LoanTerm: The term length of the loan in months
  + DTIRatio: The Debt-to-Income ratio, indicating the borrower’s debt compared to their income
  + Education: The highest level of education attained by the borrower (PhD, Master’s, Bachelor’s, High School)
  + EmploymentType: The type employment status of the borrower (Full-time, Part-time, Self-employed, Unemployed)
  + MaritalStatus: The marital status of the borrower (Single, Married, Divorced)
  + HasMortgage: Whether the borrower has a mortgage (Yes, No)
  + HasDependents: Whether the borrower has dependents (Yes, No)
  + LoanPurpose: The purpose of the loan (Home, Auto, Education, Business, Other)
  + HasCoSigner: Whether the loan has a co-signer (Yes, No)
  + Default: The binary target variable indicating whether the loan defaulted
* **Tools and Technologies Applied**:
  + Python: For data analysis, visualization, and machine learning model building.
  + Libraries:
    - Pandas & Numpy: Data manipulation and analysis.
    - Matplotlib & Seaborn: Visualization for EDA and insights presentation.
    - Scikit-learn: Building and evaluating machine learning models, such as Logistic Regression, Random Forest, and Gradient Boosting.
  + Techniques: Data cleaning, handling missing values, outlier detection, encoding categorical variables, and correlation analysis.
* **Key Insights Discovered**
  + Factors Influencing Default: Age, Income, LoanAmount, and InterestRate are significant predictors of loan default risk.
  + Younger borrowers with high loan amounts and high-interest rates show a higher likelihood of default. Non-significant Factors: Variables like Loan Purpose, Marital Status, and Employment Type showed minimal impact on default prediction, indicating they are less crucial in credit risk assessment.
* **Hypotheses based on the insights**
  + High-Interest Rates Increase Default Risk: Borrowers with loans at higher interest rates have a greater likelihood of default, likely due to the added financial burden.
  + Younger Borrowers with Larger Loans are at Higher Risk: Younger individuals with high loan amounts tend to have a higher default rate, possibly due to less financial stability or experience in debt management.
  + Stable Income and Employment Reduce Default Risk: Borrowers with consistent income and longer employment history show lower default rates, suggesting financial stability plays a significant role in repayment ability.
  + Credit scores are an important indicator of default risk. People with low credit scores (below 500) are more likely to default. This highlights the importance of credit scores in the loan underwriting process. People with high credit scores (above 700) are generally more likely to repay their loans, and therefore may be eligible for better or less risky loan terms.
* **Recommendations based on analysis results**
  + Low-Income Borrowers: For borrowers earning below 25,000, implement stricter requirements such as collateral or financial support options to reduce risk.
  + High Loan Amounts: For loans exceeding 100,000, conduct thorough financial assessments, while offering relaxed terms for smaller loans under 50,000.
  + Age-Based Adjustments: Apply more stringent checks (like collateral) for borrowers under 30. For those over 30, consider more flexible terms due to their lower risk profile.
  + Interest Rate and Credit Score: For high-interest loans (above 20%) or low credit scores, consider collateral; ease restrictions for high-credit-score borrowers