Business Problem Statement

A financial institution aims to strengthen its credit risk management process by gaining deeper insights into customer loan default behavior. Loan defaults contribute significantly to non-performing assets (NPAs), directly impacting profitability and long-term financial stability.

The organization has observed variations in default rates across different customer demographics and financial profiles, such as income levels, age groups, credit scores, and employment stability. Management wants to understand which customer attributes and behavioral factors are most strongly associated with loan default risk.

You are tasked with analyzing historical loan and customer data to address the following overarching business question:

“How can the institution leverage customer and loan data to identify high-risk segments, reduce loan defaults, and improve credit decision-making?”

Deliverables

1. **Data Preparation & Modeling (Python)**
   Clean, preprocess, and transform the raw loan dataset. Engineer relevant features to support risk analysis and ensure data quality through validation checks.

2. **Data Analysis (SQL)**
   Structure the cleaned data into analytical tables and run SQL queries to analyze default rates across customer segments based on income, age, credit score, and employment stability.

3. **Visualization & Insights (Power BI)**
   Develop an interactive dashboard to visualize default trends, high-risk customer groups, and key risk drivers, enabling stakeholders to explore insights dynamically.

4. **Report and Business Recommendations**
   Prepare a comprehensive report summarizing analytical findings, success metrics (default rate, risk segmentation), and actionable recommendations to reduce credit risk.

5. **GitHub Repository**
   Maintain a well-documented repository containing Python scripts, SQL queries, Power BI files, and project documentation for reproducibility and portfolio presentation.