Documentation: Loan Default Power BI Project

Dataset Description

The loan default dataset contains information about borrowers who have applied for loans. Each record represents an individual borrower and includes details about their **demographics**, **financial status**, **loan characteristics**, and **repayment behavior**. The purpose of analyzing this dataset is to help banks identify high-risk borrowers and make data-driven decisions on loan approvals.

Column Definitions and Purpose

The dataset includes the following key attributes:

- **LoanID** – Unique identifier for each loan.
- **Age** – Borrower's age at loan issuance.
- **Income** – Annual income of the borrower.
- **LoanAmount** – Amount requested or approved for the loan.
- **CreditScore** – Numerical representation (300–850) of creditworthiness; higher values mean lower risk.
- **MonthsEmployed** – Time in months the borrower has been with their current employer.
- **NumCreditLines** – Number of active credit accounts (cards, loans, etc.).
- **InterestRate** – Annual percentage rate (APR) applied to the loan.
- **LoanTerm** – Duration of loan repayment in months.
- **DTIRatio** – Debt-to-Income ratio (debt payments ÷ income), indicating financial stress.
- **Education** – Highest education level completed (High School, Bachelor's, Master's, etc.).
- **EmploymentType** – Employment status (Full-time, Part-time, Self-employed, etc.).
- **MaritalStatus** – Borrower's marital status (Single, Married, Divorced, etc.).
- **HasMortgage** – Yes/No indicator for an existing mortgage.
- **HasDependents** – Yes/No indicator for financial dependents (children, family).
- **LoanPurpose** – Primary reason for taking the loan (Home, Education, Business, etc.).
- **HasCoSigner** – Yes/No indicator showing if a co-signer exists for the loan.
- **Default** – Outcome column showing whether borrower defaulted (0 = No, 1 = Yes).
- **Loan Date** – Date of loan origination

Data Profiling and Type Validation in Power Query

I performed full column profiling in Power Query Editor to assess data quality across the dataset, confirming that all fields contained 100% valid values and no errors or blanks. LoanID was validated as unique, making it a suitable identifier for analysis. I then checked and confirmed that Power BI correctly assigned data types: whole numbers for Income, Loan Amount, and Credit Score; decimals for Interest Rate and DTI Ratio; text for Education, Employment Type, and Loan Purpose; Boolean values for Has Mortgage, Has Dependents, Has Co-Signer, and Default; and date type for Loan Date. This validation was important to prevent calculation errors, ensure accurate aggregations, and enable correct use of DAX functions and time-series visualizations later in Power BI.