

**Problem Statement** Financial institutions face challenges in identifying eligible applicants for home loans. The goal of this project is to explore loan application data, uncover patterns, and identify the key factors that influence loan approval decisions — helping lenders minimize risk and improve decision-making.

**Solution Overview** This analysis focuses on: - Data Cleaning: Handling missing values with appropriate imputation methods - Categorical Encoding: Label Encoding, Target Encoding - Feature Engineering: -  $\text{TotalIncome} = \text{Applicant} + \text{Coapplicant Income}$  - EMI Calculation - Affordability Check ( $\text{Able\_to\_pay\_EMI}$ ) - Visualization: Histograms, boxplots, bar plots, heatmaps - Correlation Analysis: Using Pearson & Spearman methods

**Key Insights** - Credit History is the most important factor (80%+ approval rate with good credit) - Applicants with higher combined income and lower EMI have better approval chances - Having 0 or 2 dependents increases the likelihood of approval - Most loans are taken for 25-year terms ( $\text{Loan\_Amount\_Term} = 300$  months)

**Impact** - For Financial Analysts\*: A ready EDA template for structured loan data analysis - For Learners: Understand how real-world financial data can be cleaned and analyzed step-by-step - For Developers: Base code to extend into a predictive model or Streamlit web app