Loan Default Risk Analysis

Exploratory Data Analysis (EDA) for Credit Risk Assessment

Contents

Problem Statement

- Objective: Identify high-risk loan applicants likely to default.
- Defaulted loans result in credit loss for the company.
- Goal: Analyze historical data to detect patterns and reduce financial risk.

Analysis Approach

1. Data Cleaning and Preparation
2. Univariate Analysis for individual feature behavior
3. Bivariate Analysis to assess variable relationships with defaults
4. Visual Exploration to derive insights
5. Business interpretation and actionable recommendations

Univariate Analysis Insights

- Majority of loans are Fully Paid.
- Most common loan term: 36 months.
- Interest rates concentrated around 10–15%.
- Grades 'B' and 'C' dominate the loan portfolio.
- Annual income mostly between 40K–80K.

Bivariate Analysis: Loan Grade

- Default rate increases from Grade A to G.
- Grades F and G show highest risk of default.
- Stricter loan approval recommended for lower grades.

Bivariate Analysis: Interest Rate

- Higher interest rates correlate with higher default probability.
- Applicants with high interest are riskier.
- Interest rate is a strong risk predictor.

Bivariate Analysis: Annual Income

- Lower income groups have a higher default rate.
- Wide income distribution with skewed high values.
- Minimum income thresholds can reduce risk.

Bivariate Analysis: Loan Term

- 60-month loans show significantly higher defaults than 36-month loans.
- Shorter terms are safer for lending.
- Consider higher interest or stricter criteria for long-term loans.

Key Insights & Recommendations

- Grades D to G are high risk — apply stricter filters.
- High interest and long-term loans often lead to default.
- Lower income applicants show higher risk.
- Implement risk-based pricing and improve screening for risky segments.