Loan Default Analysis

Exploratory Data Analysis and Insights

Business Understanding

- The goal is to identify factors influencing loan defaults to minimize credit loss. Insights will aid in better risk assessment and decision-making.

Dataset Overview

- - Dataset contains past loan applicants' data.
- - Key attributes: loan_status, annual_inc, grade, dti, etc.
- - Target variable: loan_status (e.g., Fully Paid, Charged-off).
- - Objective: Analyze patterns in defaults.

Data Cleaning

- - Removed columns with >40% missing values.
- - Filled numeric missing values with median.
- - Categorical missing values replaced with 'Unknown'.
- - Cleaned dataset saved for further analysis.

Univariate Analysis

- - Analyzed distributions of loan_status and annual_inc.
- - Observations:
- * Majority loans are either Fully Paid or Charged-off.
- * Annual income shows right-skewed distribution.

Bivariate Analysis

- - Relationship between variables and loan_status analyzed.
- - Key insights:
- * Lower grades (D, E) show higher default rates.
- * High debt-to-income ratio (DTI) correlates with defaults.

Correlation Analysis

- - Focused on numeric variables.
- - Mild negative correlation between annual income and default probability.
- - Stronger focus needed on highly correlated variables for prediction.

Insights and Recommendations

- - High DTI and low annual income are key risk indicators.
- - Borrowers in lower grades (D, E) are riskier.
- - Recommendations:
- * Develop stricter criteria for high-risk applicants.
- * Consider dynamic interest rates based on risk profile.

Next Steps

- - Perform statistical tests (e.g., Chi-Square, ANOVA) to validate findings.
- - Develop predictive models using machine learning.
- - Create dynamic dashboards for real-time risk assessment.