Exploratory Data Analysis (EDA) Summary Report

1. Introduction

This document presents an exploratory analysis of Geldium's dataset, aimed at evaluating data integrity, uncovering valuable insights, and identifying factors that contribute to the risk of credit default. The primary objective is to prepare the data for accurate predictive modeling and risk evaluation.

1. Dataset Summary

The dataset includes 500 customer records from Geldium, each containing essential features related to credit delinquency. It comprises both numerical and categorical data, such as earnings, credit usage, number of missed installments, and the ratio of debt to income.

Important details:

* + Total entries: 500
  + Major attributes: Age, Income, Credit Score, Credit Utilization, Missed Payments, Debt-to-Income Ratio
  + Data types:
    - Categorical: Employment Status, Credit Card Type
    - Numerical: Income, Loan Balance

1. Missing Data Evaluation

There are missing entries in crucial variables, especially in the income and loan balance fields. If left untreated, these gaps could distort model accuracy.

Observations:

* + Fields with missing data:
    - Income: 50 missing entries
    - Loan Balance: 30 missing entries
  + Planned solutions:
    - Use the median to fill missing numeric values
    - Apply AI-generated synthetic data where appropriate for Loan Balance

1. Key Insights and Risk Factors

The analysis indicates a strong link between high credit utilization and delinquency, as well as a clear risk associated with frequent missed payments.

Important insights:

* + Customers using more than 50% of their credit limit tend to be at greater risk.
  + Individuals with 3 or more missed payments within six months show a higher likelihood of defaulting.
  + Some inconsistencies were observed: high-income customers with low credit scores warrant further examination.

1. Role of AI & GenAI

Generative AI tools helped identify trends, detect missing values, and highlight risk factors. Insights from AI were validated against standard financial risk metrics to ensure reliability and accuracy.

1. Conclusion & Future Actions

The EDA revealed key delinquency indicators such as high credit usage and repeated missed payments, along with missing income and loan data. Recommended actions include applying appropriate imputation techniques and monitoring whether risk factors are consistent across customer segments.