# Exploratory Data Analysis (EDA) Summary Report

## 1. Introduction

This report provides an exploratory data analysis (EDA) of Geldium's dataset to assess data quality, identify key insights, and highlight factors that influence credit delinquency risk. The goal of this analysis is to ensure data readiness for predictive modeling and risk assessment.

## 2. Dataset Overview

The dataset contains 500 records, representing Geldium's customers with key attributes relevant to delinquency risk. It includes numerical and categorical features such as income, credit utilization, missed payments, and debt-to-income ratio.

Key dataset attributes:

- Number of records: 500
- Key variables: Age, Income, Credit Score, Credit Utilization, Missed Payments, Debt-to-Income Ratio
- Data types: Categorical (Employment Status, Credit Card Type), Numerical (Income, Loan Balance)

## 3. Missing Data Analysis

Some critical fields contain missing values, particularly in the Income and Loan Balance columns. These gaps could skew model predictions if not properly addressed.

Key missing data findings:

- Variables with missing values: Income (50 missing), Loan Balance (30 missing)
- Missing data treatment: Imputation using median values for numerical data, and AI-assisted synthetic data generation for Loan Balance where required.

## 4. Key Findings and Risk Indicators

Analysis of key risk indicators reveals that customers with high credit utilization and multiple missed payments have an increased probability of delinquency.

Key findings:

- Strong correlation between high credit utilization (>50%) and delinquency.
- Customers with 3+ missed payments in the past 6 months have a higher delinquency rate.

- Some anomalies detected where customers have high income but low credit scores, requiring further investigation.

## 5. AI & GenAI Usage

GenAI tools were used to summarize dataset trends, detect missing values, and analyze risk factors. The AI-generated insights were cross-validated against known financial risk benchmarks.

Example AI prompts used:

- 'Summarize key patterns in the dataset and identify missing values.'
- 'Analyze delinquency risk based on payment history and credit utilization.'

## 6. Conclusion & Next Steps

This exploratory data analysis (EDA) provided important insights into the quality of Geldium's dataset and key risk factors for delinquency. The analysis revealed missing financial data, clear patterns in credit behavior, and some unusual data points that need further investigation.

### *Key Findings:*

- **Missing data:** Some customers have missing income and loan balance information, which could affect predictions.
- **Delinquency risk:** Customers with high credit utilization and multiple missed payments are more likely to become delinquent.
- **Unusual data patterns:** Some high-income customers have low credit scores, which may indicate data errors or financial instability.

### *Next Steps:*

- Decide the best way to deal with missing income and loan balance values, ensuring that the chosen method does not introduce bias.
- Double-check whether high credit utilization and missed payments remain the strongest indicators of delinquency across different customer groups.
- Look into records where customers have high income but low credit scores to see if there are reporting errors or other explanations.

These findings will help Geldium refine how it assesses risk and prioritizes outreach efforts. The next steps should focus on improving data quality, verifying patterns, and preparing for further analysis.