Exploratory Data Analysis (EDA) Summary Report

# 1. Introduction

This report presents the findings from the exploratory data analysis (EDA) conducted on Geldium’s customer dataset. The goal of this analysis is to assess data quality, detect early risk indicators of delinquency, and provide actionable insights to support predictive modeling for Tata iQ’s analytics team.

# 2. Dataset Overview

Key dataset attributes:

- Number of records: 5,000 (simulated)

- Key variables: Age, Income, Credit\_Score, Credit\_Utilization, Missed\_Payments, Employment\_Status, Delinquent\_Account

- Data types: Numerical, Categorical, Binary

- Anomalies: High outliers in Income and Credit\_Utilization detected

# 3. Missing Data Analysis

Key missing data findings:

- Variables with missing values: Income (12%), Credit\_Utilization (8%), Employment\_Status (5%)

- Missing data treatment: Imputation (mean for income, median for utilization), mode for employment status

# 4. Key Findings and Risk Indicators

Key findings:

- Higher Credit\_Utilization is positively correlated with Delinquency

- Customers with 2 or more missed payments in the last 6 months show significantly higher delinquency rates

- Unexpected anomalies: Some employed customers show zero income, which requires verification

# 5. AI & GenAI Usage

Example AI prompts used:

- 'Summarize key patterns in the dataset and identify anomalies.'

- 'Suggest an imputation strategy for missing income values based on industry best practices.'

- 'Identify the top 3 variables most likely to predict delinquency.'

# 6. Conclusion & Next Steps

The dataset shows good coverage overall but contains some missing and inconsistent values that could affect model performance. Key risk indicators such as high credit utilization and missed payments have been identified as predictive features. Next steps include refining the features, validating the data further, and training delinquency prediction models.