**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 risk of credit default. The primary objective is to prepare the data for accurate modelling and risk evaluation.

# 2. Dataset Overview

The dataset includes 500 customer records Geldium’s, each contain a essential feature related to delinquency .IT comprises both Numercial and Categorial data, such as Age, earnings, credit usage, number of missed installments etc..,

- Number of records: 500

- Key variables: Age, Income, Credit Score, Credit Utilization, Missed payments, Debt-to-income Ratio.

- Data types:

* Categorial: Employment Status, Credit card type,Income etc..,
* Numercial : Age, Income, Loan, Credit Score etc..,

# 3. Missing Data Analysis

There are missing values in the crucial variables, especially in the income and loan bank field .It left untreated, these gaps could distorted model accuracy .

- Variables with missing values:

* Income:50 missing entries.
* Loan Balance :30 missing entries.

- Missing data treatment:

* Use the Median to fill the missing numerical valves.
* Apply AI -generated synthetic data where appropriate for Loan Balance.

# 4. Key Findings and Risk Indicators

The analysis indicates a strong link between High Credit Utilization and delinquency

, as well as a clear risk associated with missed payments.

-Important Insights

* Customer using more than 50% of their credit limit tend to be at greater risk.
* Individuals with 3 or more missed payments within 6 months shows a higher likelihood of defaulting.
* Some inconsistencies were observed: high income, customers with low credit scores warrant further examination.

# 5. AI & GenAI Usage

Generative AI tools supported the identification of trends, detection of missing valves, and examination of risk elements. These AI-based conclusions were compared against established financial risk metrics for validation.

-Sample AI queries:

* "Summarize data trends and highlight missing values."
* "Assess risk of default based on credit usage and payment behavior.

# 6. Conclusion & Next Steps

This EDA uncovered meaningful insights into Geldium’s dataset, highlighting missing entries, behavioral patterns tied to credit risk, and some outlier cases worth deeper analysis.

 Takeaways:

* Data gaps: Missing income and loan data could influence outcomes.
* Delinquency indicators: High credit usage and repeated missed payments are strong predictors.
* Data anomalies: Cases of high income but low credit scores need clarification.

Recommendations:

* Choose suitable imputation techniques for missing income and loan values to minimize bias.
* Confirm if key risk factors remain consistent across various customer groups.
* Investigate irregular data entries to ensure accuracy and detect potential financial instability.

These efforts will aid Geldium in refining its risk analysis processes and enchance data reliability for further modelling.

Bhuvana. J