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.

2. Dataset Overview:

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:

    o Categorical: Employment Status, Credit Card Type

    o Numerical: Income, Loan Balance

3. Missing Data Analysis:

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:

    o Income: 50 missing entries

    o Loan Balance: 30 missing entries

- Planned solutions:

    o   Use the median to fill missing numeric values

    o   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 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.

5. AI & GenAI Usage:

Generative AI tools supported the identification of trends, detection of missing values, 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.

7. 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.