**Exploratory Data Analysis (EDA) Summary**   
**Report Template**

**1. Introduction**

This document provides an exploratory analysis of Geldium’s dataset with the goal of assessing data quality, extracting meaningful insights, and identifying key factors associated with credit default risk. The main objective is to prepare the dataset effectively for reliable predictive modeling and risk assessment.

**2. Dataset Summary**

The dataset consists of 500 customer profiles from Geldium, encompassing key variables relevant to credit delinquency. It includes a mix of numerical and categorical attributes such as income, credit utilization, missed payments, and the debt-to-income ratio.  
**Key Details:**

* **Total Records:** 500
* **Main Features:** Age, Income, Credit Score, Credit Utilization, Missed Payments, Debt-to-Income Ratio
* **Data Types:**
  + **Categorical:** Employment Status, Type of Credit Card
  + **Numerical:** Income, Outstanding Loan Balance

**3. Missing Data Evaluation**

Some critical fields, notably income and loan balance, contain missing values. If not addressed, these gaps may negatively impact the accuracy of predictive models.  
**Key Observations:**

* **Missing Data Identified In:**
  + **Income:** 50 records
  + **Loan Balance:** 30 records
* **Proposed Handling Methods:**
  + Impute missing numerical values using the **median**
  + Generate synthetic values using AI techniques where suitable, particularly for **Loan Balance**

**4. Key Insights and Risk Factors**

The findings reveal a significant correlation between elevated credit utilization and delinquency, along with a notable risk tied to repeated missed payments.  
**Key Insights:**

* Customers utilizing over 50% of their available credit are more prone to default risk.
* Those with **three or more** missed payments in a six-month span exhibit a higher tendency toward credit default.
* Certain anomalies emerged, such as **high-income** individuals having **low credit scores**, suggesting the need for deeper investigation.

**5. Role of AI & GenAI**

Generative AI tools played a key role in uncovering patterns, identifying data gaps, and analyzing risk factors. The AI-generated insights were cross-verified with standard financial risk indicators to ensure accuracy and reliability.  
**Example AI Prompts Used:**

* “Identify key trends in the dataset and pinpoint missing information.”
* “Evaluate default risk based on credit utilization and payment history.”

**6. Conclusion & Future Actions**

This exploratory data analysis (EDA) revealed valuable insights within Geldium’s dataset, including data quality issues, behavioral trends linked to credit risk, and unusual cases that warrant closer inspection.  
**Key Takeaways:**

* **Missing Data:** Incomplete income and loan balance fields may impact analytical accuracy.
* **Risk Indicators:** Elevated credit utilization and frequent missed payments are strong signals of potential default.
* **Anomalies:** Instances where high-income customers exhibit low credit scores require further investigation.  
  **Recommendations:**
* Implement appropriate imputation strategies to handle missing income and loan values, reducing potential bias.
* Verify whether major risk patterns hold true across different customer segments.
* Review inconsistent data points to ensure integrity and detect any signs of financial vulnerability.

These actions will support Geldium in improving the precision of its risk evaluation framework and strengthening the quality of data used in predictive modeling.