**Exploratory Data Analysis (EDA) Report**

**1. Introduction**

This report provides an exploratory review of Geldium’s dataset, with the goal of checking data quality, extracting key insights, and identifying drivers that may increase the risk of credit default. The broader aim is to shape the dataset into a reliable foundation for predictive modeling and credit risk assessment.

**2. Dataset Overview**

The dataset contains information on 500 customers, covering a mix of demographic, financial, and credit-related variables. These attributes capture both numerical and categorical features relevant to credit behavior.

**Key details:**

* **Total records:** 500
* **Important 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 Assessment**

Some fields have incomplete values, particularly those tied to income and outstanding loan balances. If not handled properly, these gaps could bias results and reduce model reliability.

**Findings:**

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

**Proposed treatment:**

* Replace missing numeric fields (e.g., Income) with the median.
* Generate synthetic but realistic estimates for Loan Balance using AI-assisted techniques where appropriate.

**4. Core Insights and Risk Drivers**

Exploration of the data highlights strong relationships between certain behaviors and default risk. High credit utilization and repeated missed payments stand out as major red flags.

**Notable patterns:**

* Customers who use over **50% of their credit limit** are more prone to delinquency.
* Individuals with **3 or more missed installments in a six-month window** are more likely to default.
* Unexpected cases were noted, such as **high-income customers paired with unusually low credit scores**, which require closer review.

**5. Application of AI & GenAI**

Generative AI tools supported the analysis by identifying patterns, spotting missing values, and flagging possible risk elements. The AI findings were cross-checked against traditional financial risk benchmarks to ensure validity.

**Sample AI prompts used:**

* *"Identify gaps in the dataset and summarize overall patterns."*
* *"Evaluate risk of default given credit utilization and missed payments."*

**6. Conclusions & Next Steps**

This initial analysis sheds light on important characteristics of Geldium’s dataset. It uncovers missing entries, credit behavior patterns linked to delinquency, and anomalies that require further scrutiny.

**Key takeaways:**

* **Data gaps:** Missing income and loan balance values could distort analysis if not addressed.
* **Risk indicators:** Heavy credit usage and repeated late payments are strong predictors of default.
* **Unusual cases:** High earnings paired with low credit scores need to be verified.

**Recommendations:**

* Apply appropriate imputation strategies to handle missing data effectively.
* Validate whether key risk patterns remain stable across different customer segments.
* Investigate irregular data points for potential data entry errors or hidden financial risks.

By addressing these issues, Geldium can improve data reliability, refine its risk detection framework, and strengthen predictive modeling for credit evaluation.