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

# Introduction

This report summarizes the Exploratory Data Analysis (EDA) of Geldium’s customer dataset to assess data quality, identify delinquency risk indicators, and prepare for predictive modeling. Key goals include detecting missing data, anomalies, and early risk patterns.

# 2. Dataset Overview

**Key dataset attributes:**

* **Number of records:** 500
* **Key variables*:*** *Income, Credit\_Score, Credit\_Utilization, Missed\_Payments, Loan\_Balance, Debt\_to\_Income\_Ratio, Employment\_Status, Account\_Tenure, and payment history (Month\_1 to Month\_6)*
* **Data types*:*** *Numerical (Age, Credit\_Score), Categorical (Employment\_Status, Location), Binary (Delinquent\_Account)*

**Observations:**

* Missing values in Income (39), Loan\_Balance (29), Credit\_Score (2)
* Credit\_Utilization values exceed 100% (max: 1.0258)
* Inconsistent values in Employment\_Status (e.g., “EMP”, “Self-employed”)

# 3. Missing Data Analysis

**Key missing data findings:**

* **Variables with missing values:** Income, Loan\_Balance, Credit\_Score
* **Missing data treatment:**
  + Income: Impute with median – robust to outliers in financial data
  + Loan\_Balance: Impute with median – avoids skewing distribution
  + Credit\_Score: Impute with mean – only 2 missing; safe to use average

# 4. Key Findings and Risk Indicators

**Key findings:**

* **High-risk indicators:**
  + Missed\_Payments: Higher values directly correlate with delinquency risk
  + Credit\_Utilization: Over 80% suggests over-reliance on credit
  + Low Credit\_Score (<500): Associated with poor repayment behavior
  + Debt\_to\_Income\_Ratio > 40%: Indicates repayment difficulty
  + Recent Payment History: Frequent ‘Missed’ or ‘Late’ entries are red flags
* **Unexpected anomalies:**
  + Credit\_Utilization values exceed 100%
  + Inconsistent Employment\_Status values like “EMP” and “Self-employed”

# 5. AI & GenAI Usage

Generative AI tools were used to summarize the dataset, impute missing data, and detect patterns. This included identifying risk factors and forming a strategy for handling inconsistencies.

**Example AI prompts used:**

* “Summarize key patterns, outliers, and missing values in this dataset.”
* “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 is mostly complete with manageable missing values and clear high-risk indicators for delinquency. Recommended imputation strategies will prepare the data for modeling. Next steps include applying these strategies, encoding categorical variables (especially payment history), and advancing toward predictive model development.