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

# Introduction

# The primary goal of this project is to perform Exploratory Data Analysis (EDA) on customer financial data to identify risk signals and predict the likelihood of loan or credit card delinquency. Delinquency refers to customers failing to make timely payments, which is critical for financial institutions to detect early for risk management and loss prevention.

# 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 instalments, and the ratio of debt to income.

Key dataset attributes:

- Number of records: 500 total entries

- Key variables: age, income, credit\_score, credit\_Utilization, missed\_payments, debt\_to\_income ratio

- Data types:

Categorical= Employment\_status, credit\_card\_type, location, Month1..month6

Numerical= CustomerID, Age, Income, etc

# Missing Data Analysis

Identifying and addressing missing data is critical to ensuring model accuracy. This section outlines missing values in the dataset, the approach taken to handle them, and justifications for the chosen method.

Key missing data findings:

- Variables with missing values: used the formula =COUNTBLANK(col1:col\_n)

1. Income: 39 missing values

2. Loan\_Balance: 28 missing values

- Missing data treatment: Applying Mean imputation on both fields.

# 4. Key Findings and Risk Indicators

According to these analytics, it is indicated that a strong link exists between high credit utilization and delinquency, as well as a clear risk associated with frequent missed payments.

**Key findings:**

**-** Correlations observed between key variables:

* Customers with three or more missed payments within six months show a higher likelihood of defaulting.
* Individuals using more than 50% of their credit limit are expected to be at higher risk.

- Unexpected anomalies: high-income customers with low credit scores demands further examination.

# 5. AI & GenAI Usage

Generative AI tools facilitated the identification of trends, detection of missing data, and analysis of risk factors. These AI-driven inferences were contrasted with standard financial risk indicators to validate them.

Example AI prompts used:

* "Report data trends and indicate missing values."
* "Evaluate risk of default based on payment behavior and credit usage."

# 6. Conclusion & Next Steps

This EDA revealed significant findings about Geldium's dataset, including missing values, patterns of behavior related to credit risk, and instances of outliers that merit further investigation.

Takeaways:

* Missing data: Missing loan and income values may impact results.
* Delinquency predictors: High credit utilization and ongoing late payments are reliable indicators.
* Outliers: Instances of high income and low credit scores require explanation.

Recommendations:

* Select appropriate imputation methods for missing income and loan amounts to reduce bias.
* Verify whether the main risk factors are consistent throughout the different customers.
* Audit unusual entries into data to guarantee correctness and catch possible financial instability.

These initiatives will help Geldium improve its risk analysis procedures and maintain the reliability of data for subsequent modeling.