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

## 1. Introduction

The aim of this report is to examine the Geldium credit card customer dataset in order to find early indicators of customer delinquency. This will assist in developing a predictive model that classifies the customers as at risk or not at risk and aids in proactive intervention strategies.

## 2. Dataset Overview

This section summarizes the dataset, including the number of records, key variables, and data types. It also highlights any anomalies, duplicates, or inconsistencies observed during the initial review.

Key dataset attributes:

**Number of records:** 501

**Key variables:**

Customer_ID: Unique identifier

Credit_Score: Numerical score reflecting customer credit history

Credit_Utilization: Proportion of credit limit used

Missed_Payments: Count of missed credit card payments

Debt_to_Income_Ratio: Ratio between customer's debt and income

Delinquent_Account: Target variable (0 = not at risk, 1 = at risk)

**Data types:**

Categorical: Delinquent_Account, Employment_Status, Credit_Card_Type, Location

Numerical: Age, Income, Credit_Score, Credit_Utilization, Missed_Payments, Debt_to_Income_Ratio, Loan_Balance

## 3.Missing Data Analysis

Dealing with missing values guarantees quality and reliability conclusions derived from the data.

Missing variables: There were some missing values across several columns.

Variables with missing values: [Income,Credit Score,Loan Balance]

Missing data treatment: [Deletion, Imputation]

## 4.Key Findings and Risk Indicators

Examining how characteristics correlate with delinquency assisted in revealing some patt erns:

**Credit Utilization:**

Risk customers (Delinquent = 1) have higher average utilization (51%) compared to non-risk (49%).

**Credit Score:**

Interestingly, risk customers have greater average credit scores (591 vs. 575),
which shows score alone isn't a good predictor.

**Debt-to-Income Ratio:**

Risk customers indicate a slightly higher debt ratio (31% vs.
30%), which shows income might not be keeping up with borrowing.

**Missed Payments:**

Risk customers have fewer past due payments on average (2.85 vs.
2.99), illustrating that previous past due payments don't necessarily equate to future delinquency.

**Sudden anomalies:**

Young high-income customers and contrary trends in predicted risk variables (e.g., higher credit scores for delinquents) reveal non-evident risk patterns
and possible concealed variables.

## 5. AI & GenAI Usage

Generative AI software (such as ChatGPT) was employed throughout the process to help with summarization, explanation, imputation strategies, and risk factor identification.

**Example AI prompts utilized:**

"Describe the connection between credit utilization and delinquency in plain language."
"Recommend how to deal with missing income and credit score information."
"Describe trends in credit score, utilization, and missed payments for delinquent accounts."

## 6. Conclusion & Next Steps

## Summary:

The main risk signs are:

1.Credit Utilization (how much of their credit people are using)

2.Debt-to-Income Ratio (how much they owe compared to what they earn)

3.These are more useful than just looking at credit score or missed payments.

Some results were unexpected, so we can't trust just one column. We need to look at multiple factors together.

Credit score and missed payments alone do not predict risk accurately. So, it is important to look at multiple factors together.

**Next Steps:**

1. Build a simple predictive model
2. Create Visual Charts to Understand the Data using heatmap and Boxplot.