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

# 1. Introduction

This document presents an exploratory analysis of the provided delinquency prediction dataset. The goal is to assess data quality, uncover key patterns, and identify the primary factors that contribute to the risk of payment delinquency. The ultimate objective is to prepare the data for building an accurate predictive model.

# Dataset Overview

Exploratory Data Analysis (EDA) Summary Report

1. Introduction

This document presents an exploratory analysis of the provided delinquency prediction dataset. The goal is to assess data quality, uncover key patterns, and identify the primary factors that contribute to the risk of payment delinquency. The ultimate objective is to prepare the data for building an accurate predictive model.

2. Dataset Summary

The dataset contains 500 customer records, each with several features relevant to credit delinquency. It includes a mix of numerical and categorical data, covering demographics, financial status, and payment history.

**Total entries**: 500

**Key Attributes**: Age, Income, Credit\_Score, Credit\_Utilization, Missed\_Payments, Delinquent\_Account, Debt\_to\_Income\_Ratio, and a 6-month payment history.

**Data Types**:

**Categorical**: Employment\_Status, Credit\_Card\_Type, Location, and monthly payment statuses.

**Numerical**: Age, Income, Credit\_Score, Loan\_Balance, etc.

# Missing Data Analysis

The analysis identified missing entries in several critical numerical columns. Additionally, inconsistencies were found in a key categorical field, which could impact the performance of a predictive model if not addressed.

**Fields with Missing Data:**

Income: 39 missing entries

Loan\_Balance: 29 missing entries

Credit\_Score: 2 missing entries

**Data Inconsistencies:**

The Employment\_Status column contains varied representations for the same category (e.g., 'Employed', 'employed', 'EMP'). These should be standardized.

**Planned Solutions:**

Impute missing numerical values using statistical methods like mean or median imputation.

Standardize the values in the Employment\_Status column to ensure consistency.

# Key Findings and Risk Indicators

The analysis revealed a strong connection between past payment behavior and the likelihood of future delinquency.

**Top Predictors of Delinquency:**

**Missed\_Payments:** This variable has the strongest positive correlation with delinquency. A higher number of past missed payments is a direct and powerful indicator of a customer's propensity to be delinquent.

**Delinquent\_Account:** The presence of a previously delinquent account is another strong indicator of heightened credit risk.

**Credit\_Score:** This has a notable negative correlation, meaning that as the credit score decreases, the likelihood of delinquency increases. This aligns with the fundamental purpose of a credit score as a risk assessment tool.

**Outliers:** Potential outliers were detected in the Income, Loan\_Balance, and Credit\_Utilization fields. These data points may represent unique customer circumstances or data entry errors and warrant further investigation.

# AI & GenAI Usage

Generative AI was instrumental in performing this analysis. It was used to programmatically load and inspect the dataset, calculate descriptive statistics, identify and quantify missing values, generate visualizations like boxplots and heatmaps to uncover patterns and outliers, and summarize the key findings into this report.

**Sample AI Prompts Used:**

"Summarize key patterns, outliers, and missing values in this dataset."

"Identify the top 3 variables most likely to predict delinquency based on this dataset. Provide brief reasoning."

# Conclusion & Next Steps

This exploratory data analysis has provided critical insights into the dataset, highlighting data quality issues and identifying the most significant predictors of delinquency.

**Takeaways:**

**Data Gaps:** The missing Income, Loan\_Balance, and Credit\_Score data must be addressed to prevent model bias.

**Strongest Indicators:** Past payment behavior, specifically Missed\_Payments and Delinquent\_Account status, are the most powerful predictors of future delinquency.

**Data Inconsistencies:** Categorical fields require cleaning and standardization for accurate modeling.

**Recommendations:**

**Impute Missing Data:** Apply appropriate imputation techniques for the missing numerical values.

**Standardize Data:** Clean the Employment\_Status column to consolidate categories.

**Investigate Outliers:** Examine the outlier data points to determine their validity.

**Feature Selection:** Prioritize Missed\_Payments, Delinquent\_Account, and Credit\_Score as key features when building a predictive model.