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

# 1. Introduction

This report shows the **Exploratory Data Analysis (EDA)** of a customer dataset to understand their financial behavior and to help predict **delinquency risk** (whether a customer may fail to repay loans).

The purpose of this analysis is to:

* Get a clear idea of what the data looks like
* Find and fix any missing or unusual values
* Look at patterns in the data (like how income or credit score affects risk)
* Understand which columns are most useful to predict risky customers

This analysis helps prepare the data for building a good prediction model in the next steps.

# 2. Dataset Overview

This dataset contains financial and behavioral attributes of customers. It is used to predict delinquency risk based on past financial activities, credit history, and employment status. Below is a detailed explanation of each columnKey dataset attributes:

- Number of records: **500**

- Key variables:

**1.Customer\_ID** – Unique identifier for each customer (Categorical)

**2.Income** – Annual income in USD (Numerical, may contain missing values)

**3.Credit\_Score** – Credit score (typically 300–850) (Numerical)

**4.Credit\_Utilization** – Percentage of credit in use (Numerical, 0–100%)

**5.Missed\_Payments** – Missed payments in last 12 months (Numerical)

**6.Debt\_to\_Income\_Ratio** – Debt as % of income (Numerical, 0–100%) Etc.

- Data types:

-**Numerical Variables**

-**Categorical Variables**

- **Binary Variable**

# 3. Missing Data Analysis

**Here is the summary of missing data in dataset**:

| **Column** | **Missing Values** | **Percentage (%)** |
| --- | --- | --- |
| **Loan\_Balance** | 29 | 5.8% |
| **Credit\_Score** | 2 | 0.4% |
| **Income** | 1 | 0.2% |

| **Column** | **Treatment Applied** |
| --- | --- |
| Loan\_Balance | Filled with **median** |
| Credit\_Score | Filled with **mean** |
| Income | Filled with **median** |

# 4. Key Findings and Risk Indicators

Sure! Here's a **short and clear** version of your early indicators of delinquency risk:

### 🔴 **Key Indicators of Delinquency Risk**

1. **High Credit Utilization (>70%)**  
   Shows heavy reliance on borrowed money.
2. **Low Credit Score (300–600)**  
   Indicates poor credit history or past issues.
3. **Frequent Missed Payments**  
   Even 1–2 missed payments in the last year is a red flag.
4. **High Debt-to-Income Ratio (>40%)**  
   Suggests difficulty in managing debt with current income.
5. **Irregular Payment History**  
   Late or missed payments in recent months signal risk.
6. **Unstable Employment**  
   Unemployed or self-employed may have inconsistent income.
7. **Short Account Tenure (<2 years)**  
   New customers may lack repayment history.
8. **Missing or Low Income**  
   Incomplete or low income data may indicate financial instability.

📊 **Correlations Between Key Variables and Delinquency Risk**

| **🔑 Variable** | **🔗 Correlation with Delinquency Risk (Delinquent\_Account = 1)** | **📌 Insight** |
| --- | --- | --- |
| **Credit\_Utilization** | 🔺 Positive | Higher usage of available credit often signals financial stress. |
| **Missed\_Payments** | 🔺 Strong Positive | More missed payments strongly predict delinquency. |
| **Debt\_to\_Income\_Ratio** | 🔺 Positive | A high ratio suggests the borrower is over-leveraged. |
| **Credit\_Score** | 🔻 Negative | Lower credit score → higher risk of default. |
| **Income** | 🔻 Negative | Lower income generally correlates with higher risk. |
| **Employment\_Status** | 🟰 Mixed | Unemployed/Self-Employed may show higher risk depending on income stability. |
| **Account\_Tenure** | 🔻 Negative | Newer accounts are riskier; longer tenure shows trust. |
| **Age** | 🟰 Mild Negative | Younger customers may be slightly riskier, but effect varies. |
| **Month\_1 to Month\_6** | 🔺 Strong Positive (if values are 1 or 2) | Late/missed payments are direct early warning signals. |

### 🚨 **Unexpected Anomalies to Investigate**

1. **Credit Utilization > 100%**  
   Likely over-leveraged or a data error.
2. **Missing or Zero Income**  
   Suspicious if the customer has loans.
3. **Invalid Credit Scores**  
   Outside 300–850 range — likely incorrect.
4. **Negative Financial Values**  
   E.g., negative loan balance — usually not valid.
5. **Duplicate Customer IDs**  
   May indicate repeated or conflicting records.

# 5. AI & GenAI Usage

**Prompts:**

**“Give me summery of missing values”**

**“Suggest me relevant methods to fill this missing values.”**

**“Suggest me Early indicators of delinquency risk of data”**

**“Suggest me Correlations observed between key variables and Indicators”**

# 6. Conclusion & Next Steps

### ✅ **Conclusion**

The dataset has missing values in key fields like **Income**, **Loan Balance**, and **Credit Score**.  
Important features for predicting delinquency include **Credit Utilization**, **Credit Score**, and **Income**.  
Some data points are unusual, like credit use over 100% or invalid scores.

### 🧭 **Next Steps**

1. **Fix Missing Values** – Income, Loan Balance, Credit Score
2. **Check Correlations** – Between key features and delinquency
3. **Review Anomalies** – High utilization, bad scores, missing income

**--Thank You--**