# EDA Summary Report – Geldium Delinquency Prediction

## 1. Key Patterns and Anomalies

• The dataset contains 500 records with 19 columns.  
• No structural issues were found.  
• 'Income' has approximately 7.8% missing values.  
• 'Loan\_Balance' and 'Credit\_Score' have minor missing values (5.8% and 0.4%, respectively).  
• 'Account\_Tenure' includes some values as 0, which likely indicates new customers.  
• Monthly payment history columns (Month\_1 to Month\_6) require preprocessing if used for modeling.

## 2. Summary of Missing Values and How They Were Handled

|  |  |  |  |
| --- | --- | --- | --- |
| Feature | Missing % | Treatment | Reason |
| Income | 7.8% | Median Imputation | Robust to outliers |
| Credit\_Score | 0.4% | Mean Imputation | Low missing % and normally distributed |
| Loan\_Balance | 5.8% | Median Imputation | Skewed data, median is more robust |

Note: After imputation, no missing values remain in the dataset.

## 3. Risk Indicators That May Impact Delinquency Prediction

• Missed\_Payments – Strong behavioral indicator of delinquency.  
• Credit\_Utilization – High values indicate overuse of credit and financial stress.  
• Debt\_to\_Income\_Ratio – Higher values imply reduced repayment capacity.  
• Loan\_Balance – Larger balances can indicate higher default risk.  
• Account\_Tenure – Lower tenure suggests newer and possibly riskier accounts.