

Exploratory Data Analysis (EDA) Report - Geldium Delinquency Dataset

Step 1: Review the Dataset and Identify Key Insights

- Income contains missing values affecting repayment evaluation.
- Credit Utilization shows extreme high values suggesting outliers.
- Payment history (Month_1–Month_6) has missing entries causing behavior ambiguity.
- Some labeling inconsistencies exist between Missed_Payments and Delinquent_Account.

Early Indicators of Delinquency Risk:

- High Credit Utilization
- Multiple Missed Payments
- High Debt-to-Income Ratio
- Low Credit Score

Step 2: Missing Data Handling Strategy

- Income → Median Imputation
- Credit Utilization → Regression-based Imputation
- Payment History → Remove rows (<5% affected)

Step 3: Key Risk Factors

- High utilization indicates reliance on credit and increased delinquency likelihood.
- Multiple late or missed payments strongly predict future delinquency.
- Low credit score correlates with poor past repayment behavior.

Conclusion:

The dataset is suitable for modeling after cleaning missing data and resolving inconsistencies. Key risk signals such as utilization, missed payments, and credit score should be prioritized for predictive modeling.