

EDA Report: Geldium Delinquency Risk Analysis

Prepared by: Sanket Bansal
Tata iQ

Step 1: Key Insights from Initial Data Review

Notable Missing or Inconsistent Data

- The `monthly_income` field has approximately 18% missing values, particularly among younger customers or those new to the credit system.
- The `credit_utilization` field contains values exceeding 1.5, suggesting potential data entry errors or inconsistencies in reporting.
- The `last_payment_date` field shows inconsistent formats and some missing entries, primarily for customers currently classified as delinquent.

Key Anomalies

- Several entries show customers with over 10 active loans but a low debt-to-income ratio, which appears unrealistic and warrants verification.
- A subset of users displays high credit card balances but zero recorded monthly payments, indicating possible data gaps or pipeline errors.

Early Indicators of Delinquency Risk

- High credit utilization (above 90%) is strongly associated with late or missed payments.
- Missing or irregular `payment_history_3m` data is often observed in customers who have been delinquent for more than 60 days.

- A debt-to-income ratio above 0.6 frequently appears among those with recent delinquencies.

Summary of Initial Observations

The dataset contains several informative features for modeling delinquency risk, including credit utilization, payment history, and income-related ratios. However, issues related to missing income data, inconsistent formatting, and outliers must be addressed. These findings provide a foundation for refining the dataset and guiding the design of the predictive model.

Step 2: Missing Data Treatment Strategy

Feature	Identified Issue	Handling Method	Justification
monthly_income	Missing in ~18% of entries	Impute using median by age group	Median values account for demographic variation in income
credit_utilization	Contains outliers and nulls (~10%)	Cap at 1.0; impute using mean	Utilization above 100% is unrealistic; mean preserves trend
last_payment_date	Inconsistent format and missing	Standardize format; impute with mode	Ensures uniformity and reflects typical customer behavior

Step 3: Detected Risk Patterns and Factors

High-Risk Indicators

- High Credit Utilization (>90%)**
Indicates potential overextension and is a strong signal of repayment difficulty.
- Frequent Missed Payments in the Last 3 Months**
Demonstrates a behavioral trend that often precedes future delinquency.
- High Debt-to-Income Ratio (>0.6)**
Suggests financial pressure and limited repayment capacity.

4. **Recent Employment Changes**

Associated with income instability, which may increase default risk.

5. **Absence of Auto-Pay Enrollment**

Non-enrollment in auto-payment is correlated with higher likelihood of missed payments.

Additional Insights

- First-time credit users show elevated risk due to a lack of historical behavior data but also tend to respond better to early-stage interventions.
 - Older customers (above 55 years) generally exhibit lower delinquency rates despite moderate outstanding balances, possibly indicating stronger financial discipline.
-

Final Summary

The exploratory data analysis of Geldium's customer dataset reveals important risk factors for credit card delinquency. Key variables such as credit utilization, recent payment behavior, and income-debt ratios demonstrate significant predictive value. While the dataset is broadly suitable for modeling, targeted imputation and cleaning are required, particularly in income and payment history fields. These insights will inform the design of an AI-powered risk assessment tool to help Geldium prioritize outreach and reduce delinquency in a fair and transparent manner.
