2. Dataset Overview

The dataset consists of 500 customer records from Geldium, focused on analyzing credit delinquency. It include a mix of numerical and categorical features relevant to customer credit behavior.

Key dataset attributes:

* Number of entries: 500
* Key attributes:
* Age,
* Income,
* Credit score,
* Credit utilization,
* Dept to income Ratio,
* Missed payments,
* Loan balance,
* Account tenure
  + Data types:
* Numerical: Income, Loan balance, Credit score
* Categorical: Customer ID, Employment status, Credit card type

# 3. Missing Data Analysis

A preliminary assessment reveals that certain crucial fields contain missing values, especially those critical for financial modeling

Missing Data Summary:

* Income: 50 entries missing
* Loan Balnce:30 entries missing

Proposed Treatment:

* Numerical Fields (e.g.. Income): Fill using median imputation to reduce the impact of outliers.
* Loan Balance: Consider AI-generated synthetic data to enhance completeness and maintain distribution integrity.

# 4. Key Insights and Risk Factors:

Exploratory analysis revealed notable trends and indicators that contributes to credit delinquency risk.

Significant findings:

* Customers with credit utilization over 50% are more likely to default.
* Those with three or more missed payments within six months show increased default risk.
* Some anomalies identified : high-income individuals displaying low credit scores, which requires further investigation.

# 5.Role of AI & GenAI

AI and Gen AI tools were instrumental in:

* Identifying patterns and missing data.
* Highlighting high-risk customer segments.
* Supporting exploratory data analysis (EDA)

Example AI prompts used:

* “Summarize data trends and highlight missing values”
* “Asses default risk based on credit usage and payment history”

These insights were cross-validated using standard financial risk metrics.

# 6. Conclusion & Recommended Next Steps

The exploratory data analysis provided valuable insights into customer credit behavior and flagged areas needing deeper examination.

Key Takeaways:

* Data Quality: Income and loan balance fields need careful imputation.
* Risk Signals: High Credit usage and missed payments are key indicators of default risk.
* Outliers: High-income ,low credit-score cases demand further analysis.

Next Steps:

* Apply robust imputation method to minimize modeling bias.
* Evaluate consistency of risk indicators across customer segments.
* Investigate anomalous records for potential data entry errors or hidden financial issues.

This analysis will help Geldium improve credit risk evaluation ,enhance data completeness, and prepare for predictive modeling with greater confidence.