**Exploratory Data Analysis (EDA) Summary**

**1. Introduction**

The Geldium's Collections team wants to predict repayment risk and manages customer delinquency using AI driven insights.

The main goal of this are improve efficiency and reduces the number of missing payments using better decision making.

**2. Dataset Overview**

Key dataset attributes:

- Number of records: 500

- Key variables:

Variable Description

Customer\_ID Unique Id for each customer

Age Customer age

Income Annual Income

Credit\_Score Credit Rating(Lower=Worst)

Credit\_Utilization %of credit that is used out of credit

Missed\_Payment Total number of payments that are missed

Delinquet if target is 1 then delinquent else not

Debt\_To\_Income Ratio of loan debt to income

Repayment\_History for month 1-6

- Data types:

Variable Data Types

Customer\_ID Categorial

Age Numerical

Income Numerical

Credit\_Score Numerical

Credit\_Utilization Numerical

Missed\_Payment Integer

Delinquet Binary

Debt\_To\_Income Numerical

Repayment\_History Categorical

**3. Missing Data Analysis**

Identifying and addressing missing data is critical to ensuring model accuracy. This section outlines missing values in the dataset, the approach taken to handle them, and justifications for the chosen method.

Key missing data findings:

- Variables with missing values: Loan\_Balance, Income

- Missing data treatment: Deletion for Loan\_Balance and Income can be filled using mean,median,mode concept.

**4. Key Findings and Risk Indicators**

This section identifies trends and patterns that may indicate risk factors for delinquency. Feature relationships and statistical correlations are explored to uncover insights relevant to predictive modeling.

Key findings:

- Correlations observed between key variables:

1.High credit utilization increases likelihood of missing payments.

2.Lower credit score are higher chances of delinquency.

3.Lower income correlate higher risk, but not always.

- Unexpected anomalies: High credit Utilization, low credit scores, income is 0 or low,suspecious repayment patterns, and missing values.

**5. AI & GenAI Usage**

Generative AI tools were used to summarize the dataset, impute missing data, and detect patterns. This section documents AI-generated insights and the prompts used to obtain results.

Example AI prompts used:

- Summarize key patterns in the dataset and identify anomalies.

**6. Conclusion & Next Steps**

The analysis are as follows:

1.High credit utilization, lower credit score, and multiple missed payments,lower income are the most significant predictors.

2.Some data like missing values, inconsistent categorial entries need more attention.

3.Logistic regression or decision tree is used for early risk prediction.

**Next steps:**

Recheck the data

Model development

Targeted collection strategy

Measure business impact