**1. Introduction**

The purpose of this report is to explore patterns, detect anomalies, and prepare the data for a predictive model aimed at identifying **delinquency risks** in customer financial behaviour.

**2. Dataset Overview**

* **Number of records:** 500
* **Number of features:** 19
* **Key variables:**
  + Age, Income, Credit\_Score, Missed\_Payments
  + Delinquent\_Account (likely target)
  + Monthly status columns: Month\_1 to Month\_6
* **Data types:**
  + Numerical: int64, float64
  + Categorical: object

**Issues Detected:**

* Missing values in:
  + Income: 39 missing
  + Credit\_Score: 2 missing
  + Loan\_Balance: 29 missing

**3. Missing Data Analysis**

* **Variables with missing values:**
  + Income, Credit\_Score, Loan\_Balance
* **Suggested Treatments:**
  + **Income:** Impute with median or use predictive imputation (e.g., based on Employment\_Status)
  + **Credit\_Score & Loan\_Balance:** Median imputation likely acceptable

**4. Key Findings and Risk Indicators**

**a) Correlation Matrix (Numerical Variables)**

* Most correlations are weak or close to zero.
* Slight inverse relationships:
  + Missed\_Payments & Account\_Tenure: -0.097
  + Missed\_Payments & Debt\_to\_Income\_Ratio: 0.00015 (very weak)
* Delinquent\_Account has low correlation with all variables. A stronger model may require:
  + Feature engineering
  + Combining monthly trends

**b) Numerical Distributions**

* **Age:** Fairly uniform across adult ranges
* **Income:** Slight right skew
* **Credit\_Score:** Mostly between 300–800
* **Delinquent\_Account:** Highly imbalanced (most are 0)

**5. AI & GenAI Usage**

AI tools were used to:

* Summarize patterns across all numerical and categorical data
* Visualize missing data and correlations
* Suggest imputation strategies based on domain practices

**Example Prompts:**

* *"Summarize key trends and identify outliers in customer financial data"*
* *"Suggest imputation techniques for missing income based on similar employment status"*

**6. Conclusion & Next Steps**

**Key Takeaways:**

* Missing data is manageable
* Weak correlations suggest need for:
  + Feature interactions (e.g., combining monthly payments into trends)
  + Model-based feature selection
* Delinquency target is highly imbalanced

**Next Steps:**

* Impute missing values
* Create aggregate features (e.g., % missed payments across 6 months)
* Encode categorical variables
* Balance the dataset (e.g., SMOTE, class weighting)
* Train classification models (e.g., Logistic Regression, Random Forest)