**Delinquency Risk Analysis Report**

**Introduction to Delinquency Prediction Dataset**

Understanding delinquency risk is crucial in financial decision-making, helping lenders assess the likelihood of customers failing to meet financial obligations. The **Delinquency Prediction Dataset** serves as a valuable resource for identifying patterns and risk factors associated with loan repayment behaviour.

**Key Influencing Factors**

Based on the dataset analysis, the following variables have the most significant influence on delinquency risk:

1. **Credit Score** – A low credit score is a strong predictor of financial instability and inability to meet payments.
2. **Credit Utilization** – High credit usage indicates financial stress, increasing the likelihood of delinquency.
3. **Missed Payments** – Past missed payments are one of the strongest indicators of future delinquency risk.
4. **Debt-to-Income Ratio** – A higher ratio suggests financial burden, making delinquency more likely.
5. **Employment Status** – Unstable employment or unemployment can lead to difficulty in making timely payments.

**Variables with Missing Data**

The dataset contains missing values in the following variables:

* **Loan Balance**
* **Debt-to-Income Ratio**
* **Delinquent Account**
* **Employment Status**
* **Account Tenure**

**Potential Issues Due to Missing Data**

1. **Bias in Predictions** – Missing values in crucial financial variables such as Loan Balance and Debt-to-Income Ratio can lead to inaccurate predictions of delinquency risk.
2. **Reduced Model Accuracy** – Incomplete data negatively impacts the training of models, potentially leading to weak predictive power.
3. **Data Imbalance** – If missing values are skewed toward specific categories (e.g., unemployed individuals), it can lead to biased risk assessments.
4. **Inconsistent Insights** – Employment Status gaps can make it difficult to establish a clear connection between financial stability and delinquency.
5. **Difficulty in Credit Decisions** – Missing Delinquent Account data means past financial behavior may not be fully accounted for in risk evaluations.

**Imputation Strategy for Variable with missing value**

| **Variable** | **Suggested Imputation Method** | **Reasoning** |
| --- | --- | --- |
| **Loan Balance** | Median Imputation | Loan balances may have extreme values, so using the median helps avoid skewed results while preserving realistic values. |
| **Debt-to-Income Ratio** | Mean Imputation | Debt-to-income ratio generally follows a normal distribution, making mean imputation a suitable approach for maintaining stability in risk assessment. |
| **Delinquent Account** | Mode Imputation | Since this is a categorical variable (0 or 1), replacing missing values with the most frequent value prevents unwanted bias in delinquency prediction. |
| **Employment Status** | Predictive Imputation | Employment status impacts financial stability. Using a classification model (e.g., logistic regression) to estimate missing values based on other financial indicators ensures logical imputation. |
| **Account Tenure** | Median Imputation | Account tenure values may vary widely. Using the median prevents distortion due to extreme values while ensuring consistency in stability analysis. |

**High-Risk Indicators for Delinquency Prediction**

Here are key **high-risk indicators** that significantly influence delinquency prediction, along with explanations and insights:

**1. Credit Score (<400)**

* **Explanation:** A very low credit score indicates poor past financial behavior and a high likelihood of default.
* **Impact on Prediction:** Borrowers with **low credit scores** often exhibit high delinquency rates due to previous missed payments or excessive debt.

**2. High Credit Utilization (>70%)**

* **Explanation:** Using a large percentage of available credit suggests financial strain and difficulty in managing debt.
* **Impact on Prediction:** High **credit utilization** may indicate increased reliance on borrowing, signaling risk of repayment issues.

**3. High Debt-to-Income Ratio (>50%)**

* **Explanation:** A high debt-to-income ratio suggests excessive financial obligations relative to earnings, leading to delinquency risk.
* **Impact on Prediction:** Borrowers **struggling with debt repayment** may be unable to meet financial obligations, increasing default rates.

**4. Multiple Missed Payments (≥3 in past 6 months)**

* **Explanation:** Frequent missed payments signal ongoing financial difficulties and a pattern of late repayment.
* **Impact on Prediction:** **Past delinquency** is a strong predictor of future default, making this an essential factor in assessing risk.