Business Summary Report: Predictive Insights for Collections Strategy

# 1. Summary of Predictive Insights

|  |  |  |  |
| --- | --- | --- | --- |
| Key Insight | Customer  Segment | Influencing Variables | Potential Impact |
| A small group of customers are at high risk for missing payments even if their credit scores seem okay. This goes against the common belief that credit scores are the only thing that matters. | Customers who show a high chance of missing payments, like those on the top of our risk list. (e.g., CUST0159, CUST0478). | How much of their available credit they are using, and their past history of missed payments. | The collections team can focus on a new group of at-risk people, instead of just those with low credit scores. This helps find hidden risks and be more proactive. |
| The rate of missed payments is higher in some cities than others. | Customers residing in Los Angeles (19%) and Houston (17%) have a higher base risk of delinquency than those in New York (12.5%) and Chicago (13%). | Where a customer lives. | The collections team can give more attention to customers in higher-risk areas, making their efforts more efficient and targeted. |
| People who have missed a few payments are just as likely to miss more as those who have missed many. | Customers who have a moderate number of missed payments are also flagged as high-risk. | The combination of a person's financial details, not just a simple count of missed payments. | The team can find a wider range of at-risk customers, not just the most obvious ones. This allows them to step in earlier to prevent more problems. |
| We can pinpoint a small, specific group of customers who are most likely to miss payments. | The top 10 to 20 customers on our risk list. | A combination of how much credit they use, their debt-to-income ratio, and other financial details. | The collections team can create a priority list for who to contact, making sure they use their time and resources on the people who need it most. |
| Recent payment behavior is a strong clue for future risk. | Customers who have recently been "Late" or "Missed" payments in the last few months. | The pattern of recent payments. | The collections team can set up an alert system to find customers who are starting to have trouble, so they can offer help before the situation gets worse. |

# 2. Recommendation Framework

* **Restated Insight:** A small group of customers, even with seemingly acceptable credit scores, are identified as being at high risk for delinquency, which challenges our standard risk assumptions.
* **Proposed Recommendation:**
* **Specific:** Create a prioritized collections list based on the top 20 customers with the highest predicted delinquency probability.
* **Measurable:** The success will be measured by a 15% reduction in the number of these customers becoming 90+ days delinquent over the next quarter.
* **Actionable:** The collections team will use this new list for proactive contact, rather than relying solely on traditional credit score metrics.
* **Relevant:** This directly addresses the high-risk customer segment identified by the data and focuses resources on the most likely defaulters.
* **Time-bound:** The new collections strategy will be implemented within one week, with performance measured at the end of the next quarter.
* **Justification and Business Rationale:** By shifting from a reactive collections strategy based on old metrics to a proactive one using data-driven insights, we can prevent a significant number of defaults. This targeted approach is more efficient and will likely lead to reduced losses and improved portfolio health.

# 3. Ethical and Responsible AI Considerations

**Fairness, Transparency, and Impact of the Model:**

* **Potential for bias or unfair treatment:** The initial data analysis revealed a risk of bias, with delinquency rates varying across locations. For example, customers in Los Angeles have a higher delinquency rate than those in New York. While our model failed to make any positive predictions, if it were to function, it could disproportionately flag customers from certain locations, leading to unfair treatment.
* **Explainability:** The choice of a Logistic Regression model supports transparency. Its predictions are easily explainable through model coefficients, which allows us to clearly communicate why a customer was identified as high-risk.
* **Responsible financial decision-making:** The recommendation for a proactive, data-driven collections strategy supports responsible financial management. It focuses on early intervention to help customers avoid further financial distress.
* **Other ethical principles considered:** Our approach prioritizes transparency by using an explainable model and accountability by recommending a human review and approval (human-in-the-loop) process for a final review of the model's predictions. Data privacy is also a critical consideration, as all customer information is treated as confidential.

**Based on the logistic regression model, here is a list of the top 10 customers most likely to miss payments in the future, ranked by their predicted probability:**

| Customer\_ID | Predicted Delinquency Probability |
| --- | --- |
| CUST0159 | 0.329327 |
| CUST0478 | 0.326517 |
| CUST0150 | 0.322289 |
| CUST0112 | 0.321528 |
| CUST0167 | 0.318845 |
| CUST0192 | 0.315742 |
| CUST0171 | 0.314114 |
| CUST0361 | 0.314011 |
| CUST0293 | 0.312120 |
| CUST0109 | 0.308835 |