Analysis of Paperless Banking Enrollment at Sandhills Bank

Saakshi Agarwal

This study analyzes the relationship between customer balance and paperless banking enrollment at Sandhills Bank. The bank aims to increase the adoption of paperless banking and wants to predict which customers are most likely to enroll. The dataset consists of 1,000 observations, and balance is used as the independent variable, as financial engagement is expected to influence enrollment decisions.

The mean balance ($1,122) and median balance ($1,045) for enrolled customers are higher than for non enrolled customers (mean: $510.7, median: $507), as seen in Table 1. Figure 1 shows that the interquartile range (IQR) is much larger for enrolled customers, indicating a wider spending distribution. It can also be observed that some people who are not enrolled have a high balance. The minimum balance for enrolled customers ($365) is more than double that of non-enrolled customers ($76), suggesting that customers with lower balances are less likely to transition to paperless banking.

Table 1: Descriptive analysis for Balance (Dollars) based on being Enrolled

Enroll	Min (Dollar)	Q1 (Dollar)	Median (Dollar)	Mean (Dollar)	Q3 (Dollar)	Max (Dollar)
Yes	365	898	1045	1122	1307	2726
No	76	365	507	510.7	638.5	2498

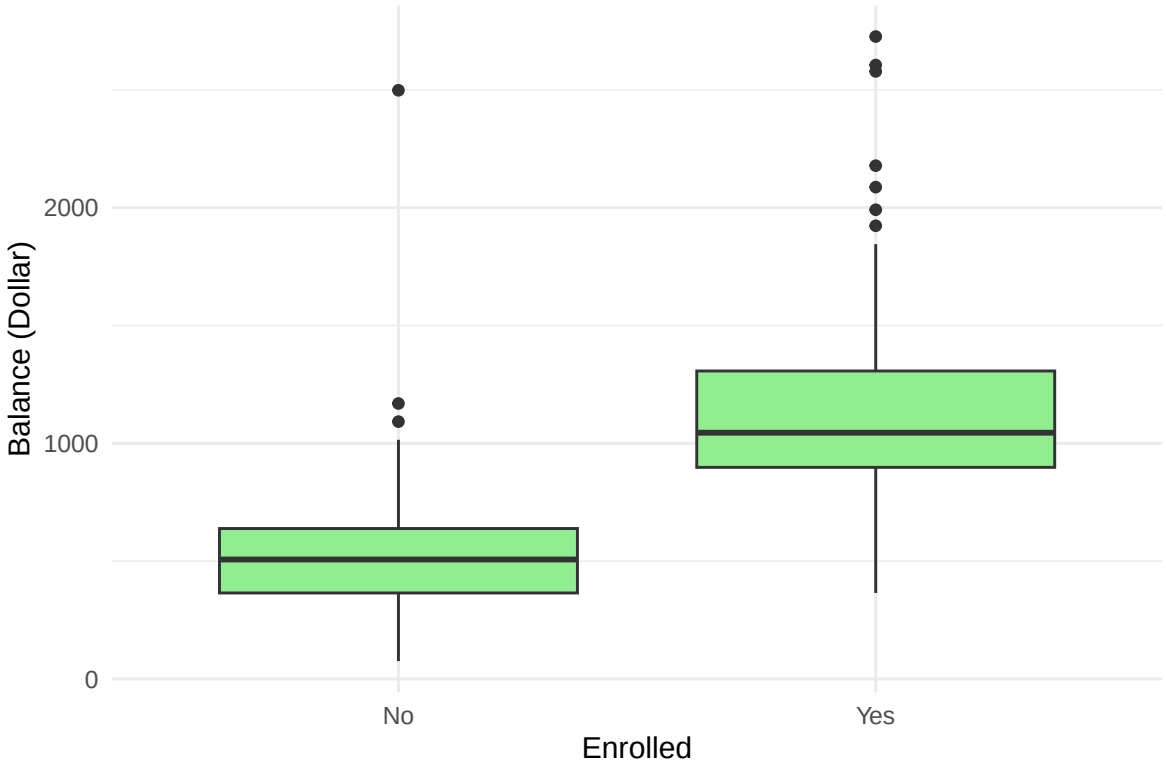

Figure 1: Boxplot for Balance based on Enrollement status

A logistic regression model is used to predict whether a customer enrolls in paperless banking. As seen in Table 2, for every $1 increase in balance, the odds of enrolling are 1.009 times higher than the odds of not being enrolled. In other words, the odds of being enrolled increase by 0.9%. As the p-value is below 0.001, it confirms that balance is statistically significant for predicting enrollment. For example, customer with balance of $1,500 has 99.8% probability of enrolling. @fid-sigmoid shows that

the probability of enrollment remains low for small balances but increases sharply around $1,000. It approaches a probability of 1 for balances nearly above $1,500, indicating that high-balance customers are almost certain to enroll.

Table 2: Summary of Logistic Regression Model

Coefficient	Estimate	Odds Ratio	Std_Error	P-Value
y = Intercept	-7.98	0.000342 (e^-7.98)	0.713	4.42e-29 (<0.001)
x1 = Balance (Dollar)	0.00944	1.009485 (e^0.00944)	0.00	8.74e-27 (<0.001)

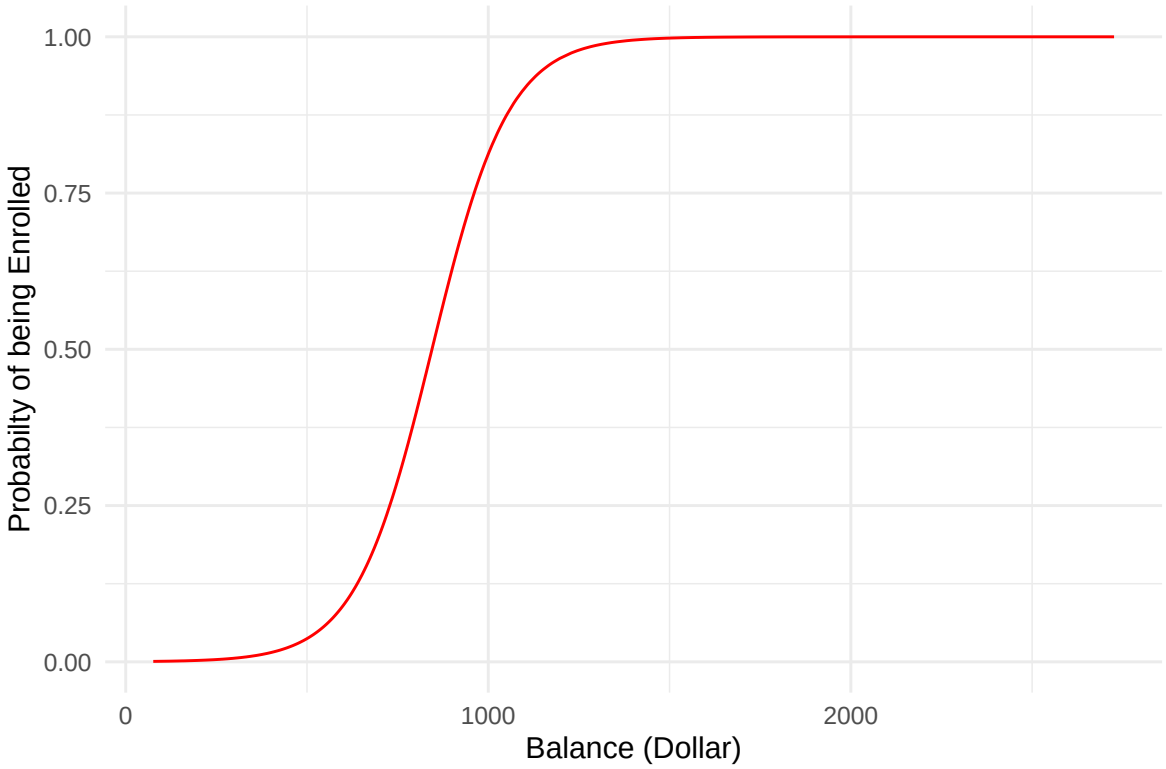

Figure 2: Probabilty of Enrolling based On Balance

The model correctly predicted 328 customers who would not enroll (True Negative) and 133 who would enroll (True Positive). However, 28 enrolled customers were misclassified as non-enrolled (False Negative), and 12 non-enrolled customers were incorrectly classified as enrolled (False Positive) [Table 3]. The model's overall accuracy is 92%, suggesting that the model correctly predicts most cases. It correctly identifies 96.5% of enrolled customers (sensitivity) and 82.6% of non-enrolled customers (specificity). Of all customers predicted to enroll, 91.7% actually are enrolled (ppv), while of all predicted non-enrolled, 92.1% did not enroll (npv). The ROC AUC score is 0.973, and the curve is well above the diagonal Figure 3, confirming that the model effectively distinguishes between enrolled and non-enrolled customers.

Table 3: Confusion Matrix for Logistic Regression Model

	Actual: No	Actual: Yes
Prediction: No	328	28
Prediction: Yes	12	133

Table 4: Summary of Classification Metric

Metric	Estimate
Accuracy	0.920
Sensitivity	0.965
Specificity	0.826
Positive Predictive Value (ppv)	0.921
Negative Predictive Value (npv)	0.917

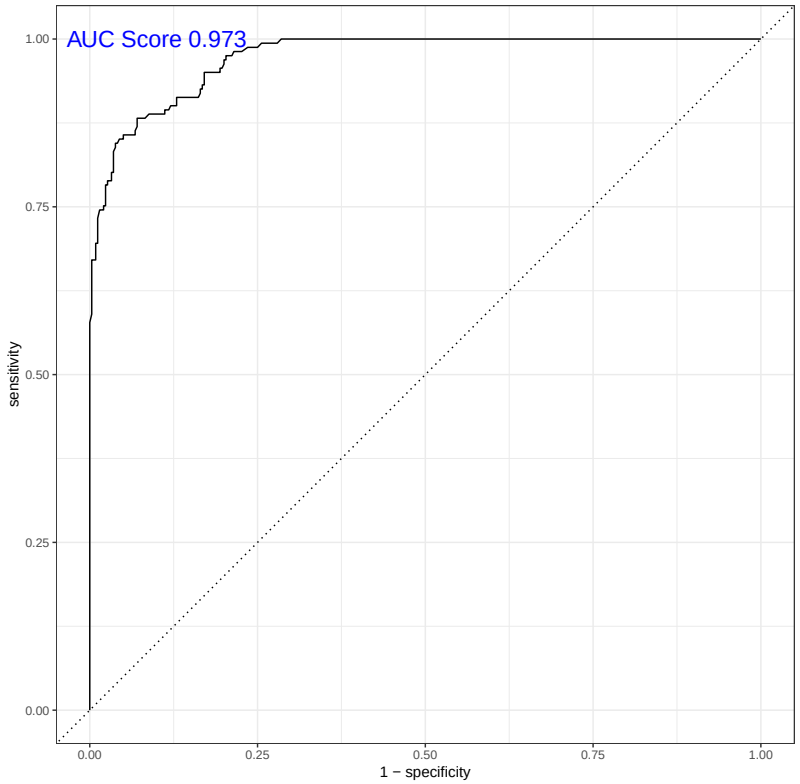

Figure 3: ROC Curve for Model Accuracy

For Sandhills Bank, this analysis provides a clear advantage in targeting non-paperless customers with high balances for promotional efforts. By focusing marketing campaigns on the 200 non-paperless customers, especially those with balances above $1000, the bank can increase adoption rates, improve efficiency, and enhance customer engagement.