CREDIT CARD BEHAVIOUR SCORE PREDICTION REPORT

USING CLASSIFICATION AND RISK-BASED TECHNIQUES

Presented By

ZAWAR ISHITA AASHISH

(23118082)

OVERVIEW OF APPROACH AND MODELING STRATEGY:

The goal of this project is to develop a forward-looking Behaviour Score model for Bank A, to identify credit card customers who are likely to default in the next month. This helps the bank take proactive actions like reducing credit exposure, activating early warnings, and prioritizing risk management.

Modeling Strategy:

1.Data Understanding and Preprocessing
2.Data exploration (EDA) to uncover behavioral trends
3.Feature Engineering with Financial Logic
4.Class imbalance handling using SMOTE
5.Model selection & tuning across Logistic Regression, XGBoost, LightGBM
6.Evaluation using risk-aware metrics, especially F2 score
7. Model Explainability (SHAP)
8.Final prediction on an unseen validation set

EDA FINDINGS AND VISUALIZATIONS:

Univariate Analysis:

Categorical Features:

Sex:

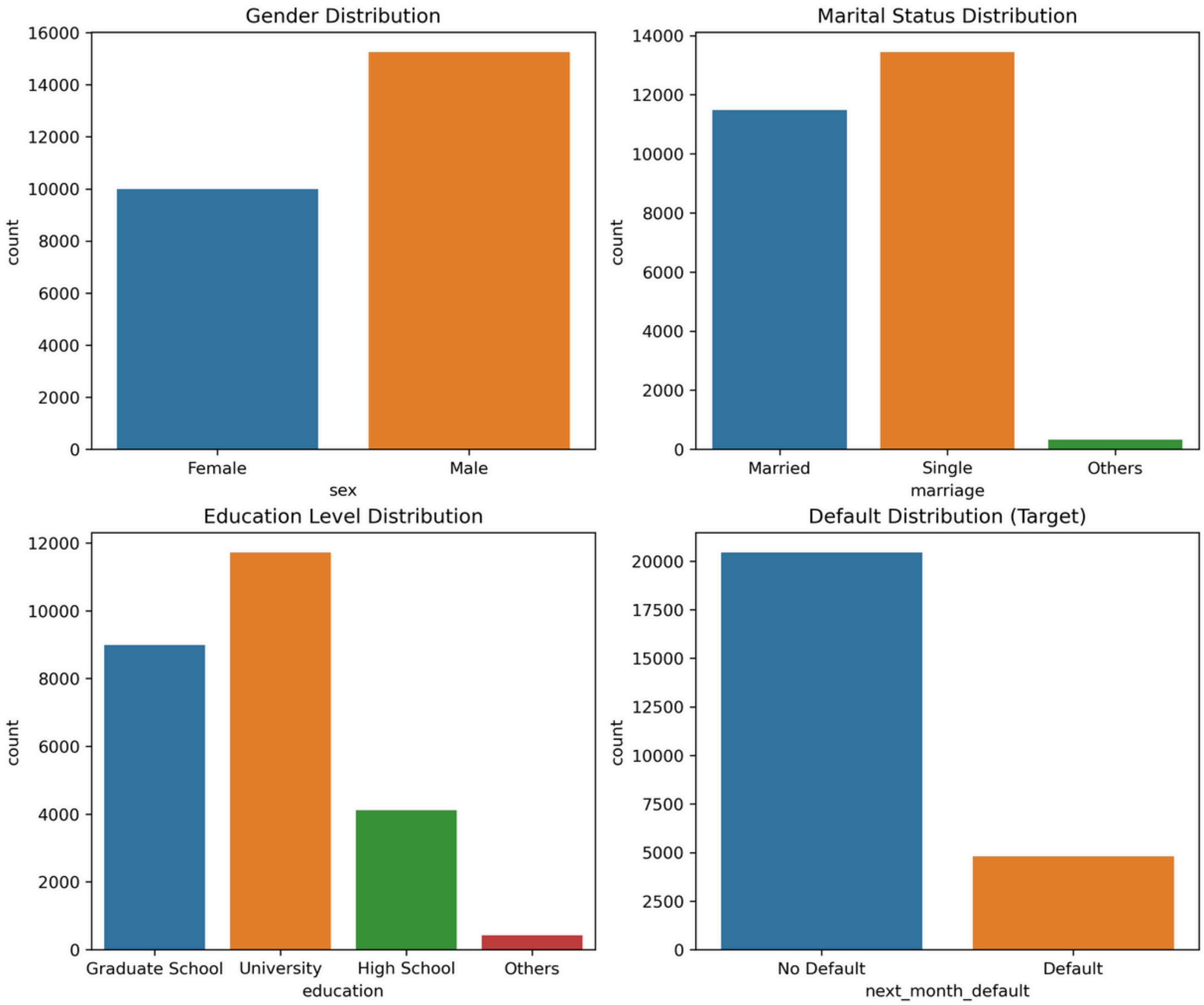
- Most customers are male (~60%).
- Default rate is slightly higher among females (20.9%) than males (17.8%).

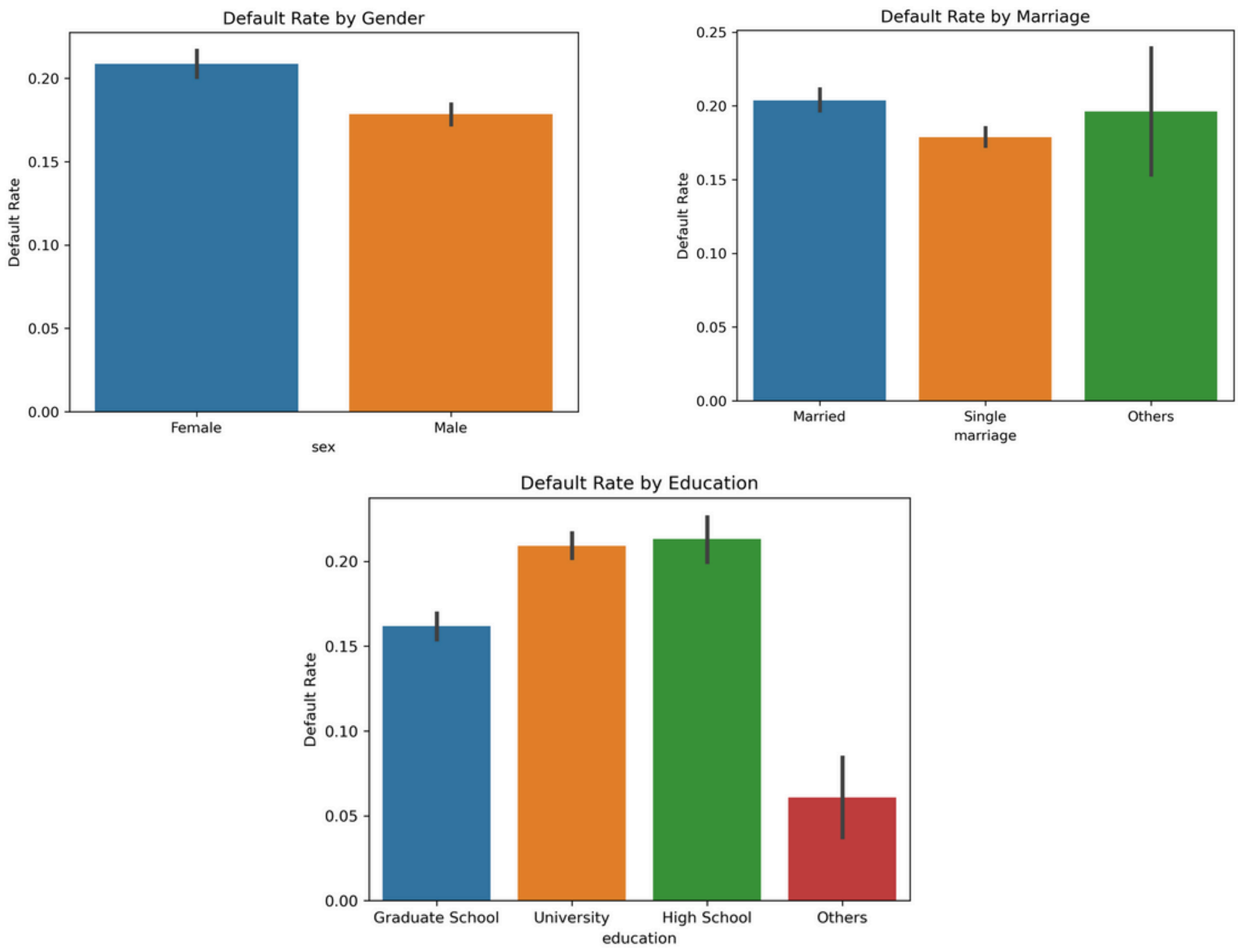
Marriage:

- Majority were single followed by married individuals.
- Default rate more among married individuals.

Education:

- Most customers have university or graduate degrees.
- Default rate more among high school and university students.


Gender Distribution
count
16000
14000
12000
10000
8000
6000
4000
2000
0
Female
Male
sex
Marital Status Distribution
count
14000
12000
10000
8000
6000
4000
2000
0
Married
Single
Others
marriage
Education Level Distribution
count
12000
10000
8000
6000
4000
2000
0
Graduate School
University
High School
Others
education
Default Distribution (Target)
count
20000
17500
15000
12500
10000
7500
5000
2500
0
No Default
Default
next_month_default


Default Rate by Gender
Default Rate
0.20
0.15
0.10
0.05
0.00
Female
Male
sex
Default Rate by Marriage
Default Rate
0.25
0.20
0.15
0.10
0.05
0.00
Married
Single
Others
marriage
Default Rate by Education
Default Rate
0.20
0.15
0.10
0.05
0.00
Graduate School
University
High School
Others
education

Numerical Features:

Credit Limit (LIMIT_BAL) :

- Right skewed distribution.
- Many customers have moderate limits, a few have very high limits.

Age:

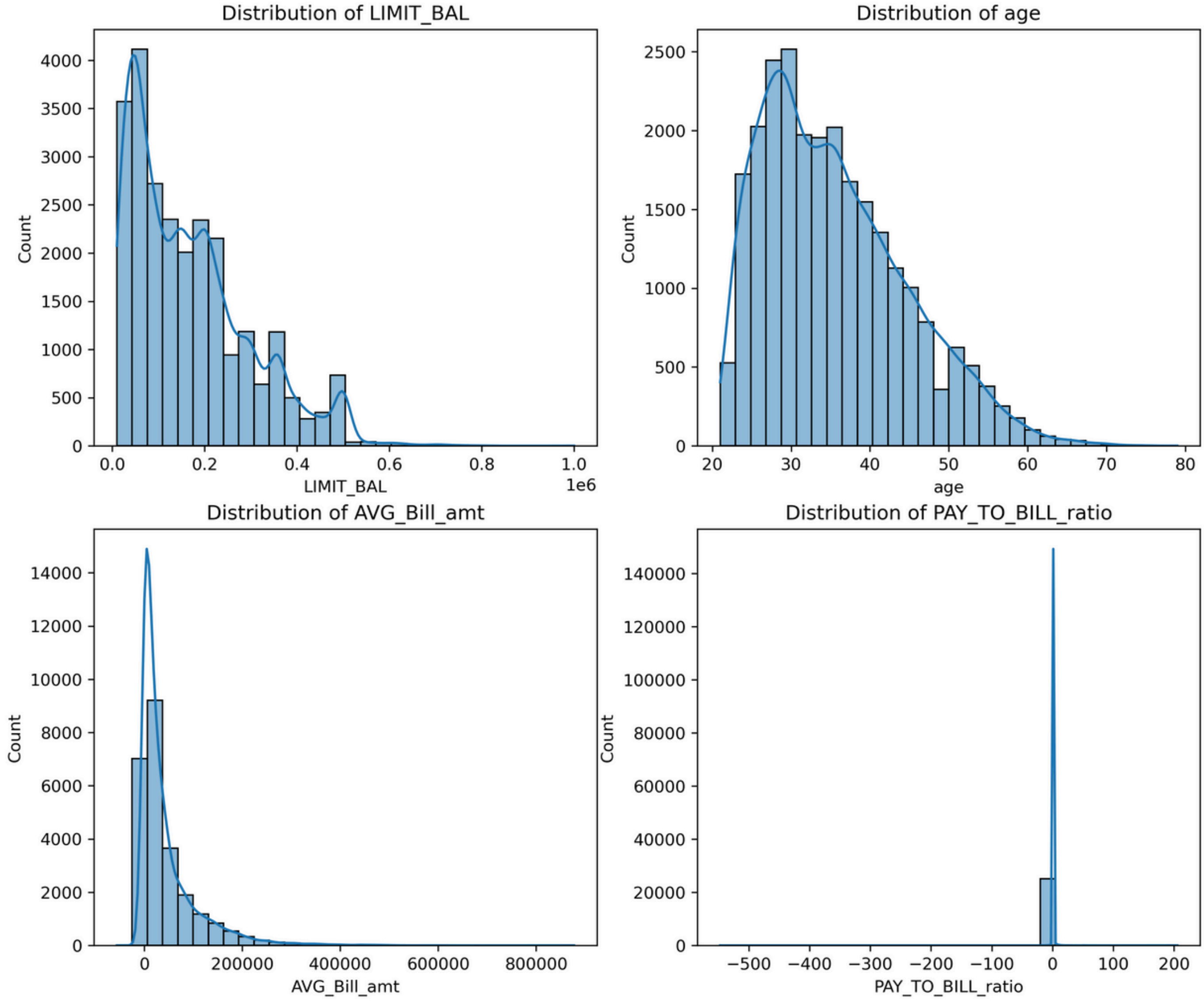
- Mostly customers are between 30–50 years.
- Slight increase in default rates among customers around 30 and below.

AVG_Bill_amt :

- Most customers have modest average bill amounts, with the peak around ₹0–₹50,000.
- A long tail indicates that a small number of customers have very high average bills (₹200,000+), possibly high spenders or heavy credit users.
- There are some customers with AVG_Bill_amt < 0, which can happen if a customer overpays.

PAY_TO_BILL_ratio:

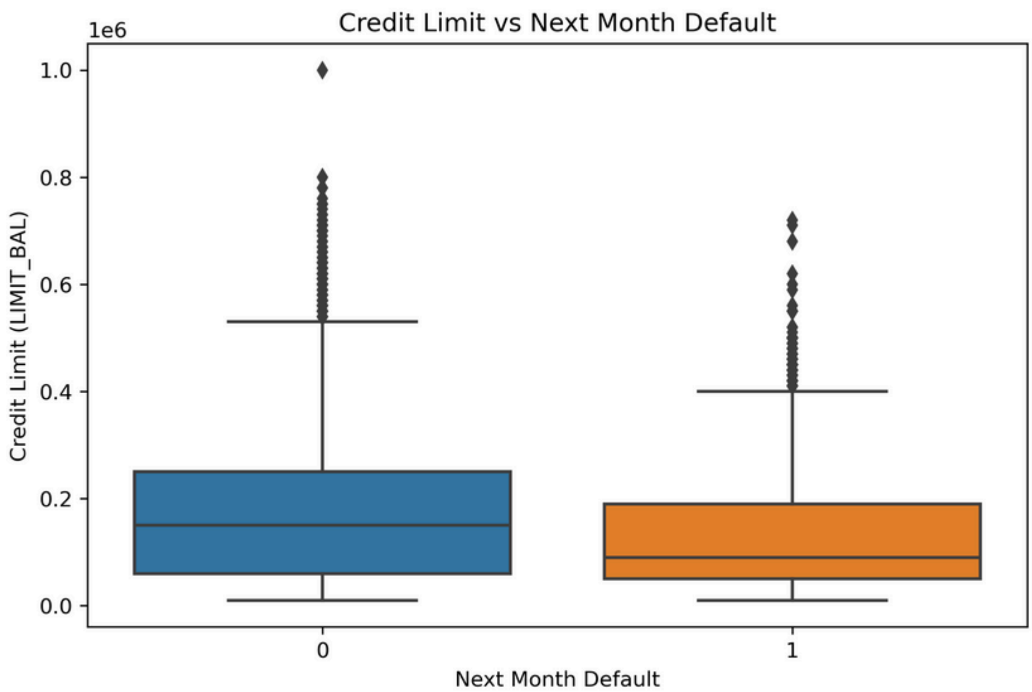
- The bulk of the values are tightly packed around 0–1, meaning most customers pay either the full bill or a partial amount.
- Some values are greater than 1, indicating overpayment, while others are below 0 — which may result from credit reversals, refunds, or data errors.



Bivariate Analysis:

LIMIT_BAL vs Next Month Default:

- The median credit limit is higher for non-defaulters than defaulters.
- Defaulters tend to have lower credit limits, with most values concentrated below ₹150,000.
- Non-defaulters have a wider spread and higher outliers, showing the bank assigns larger limits to more trustworthy customers.
- This suggests that customers with lower limits are more likely to default, possibly due to financial constraints or low creditworthiness.



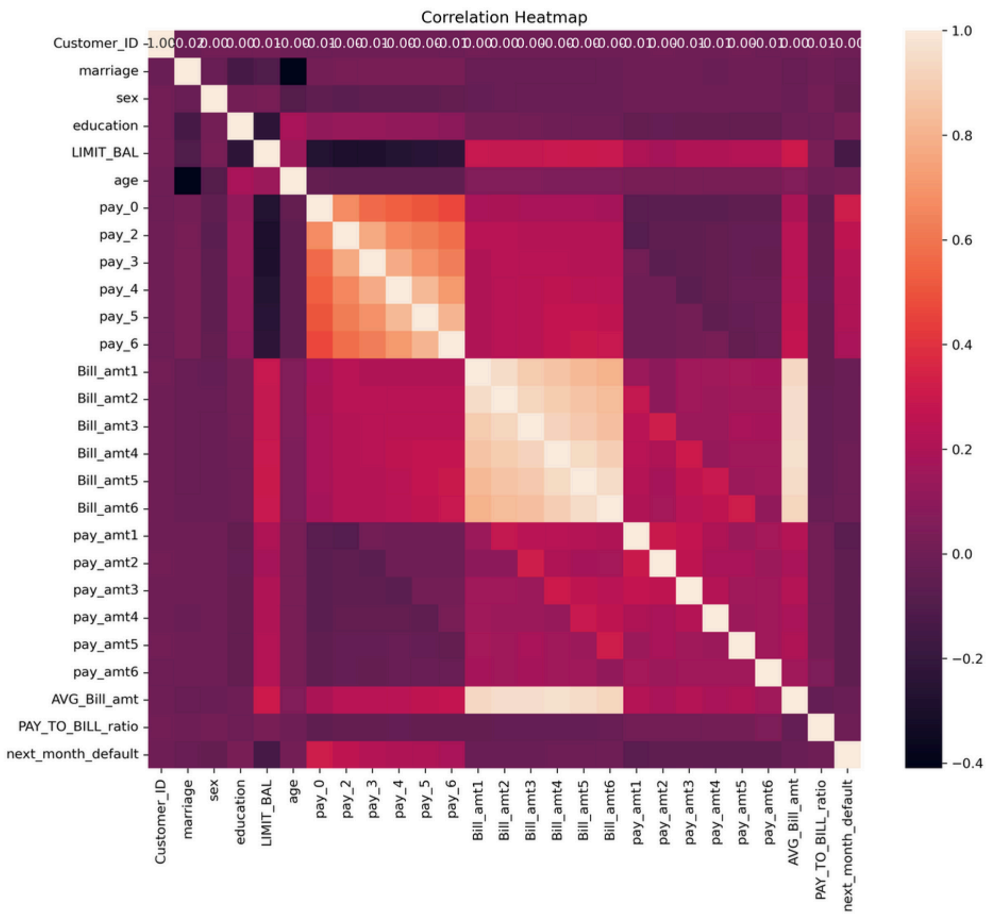
Correlation Heatmap:

Strong positive correlation observed between:

- PAY_0 to PAY_6: showing consistent payment delay behavior.
- Bill_amt1 to Bill_amt6 : monthly bills are closely related over time

next_month_default shows:

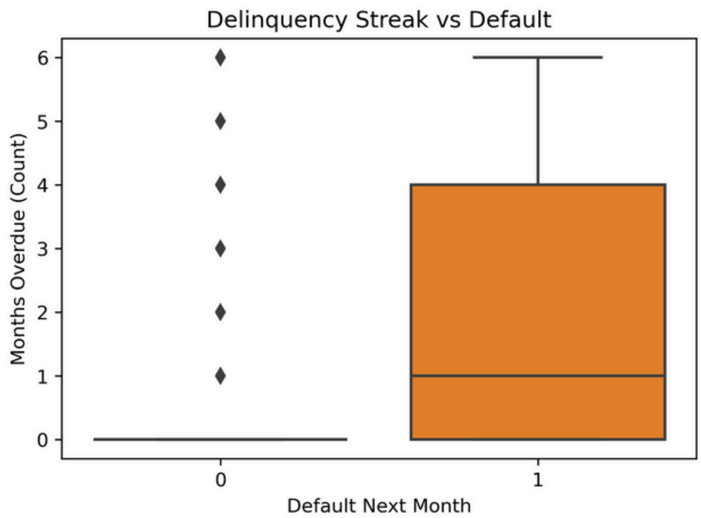
- Moderate positive correlation with pay_0,pay_6: Indicates that delayed payments are associated with higher default risk
- Weak negative correlation with LIMIT_BAL, PAY_TO_BILL_ratio: Lower credit limits and lower repayment ratios slightly increase default chance



FINANCIAL INSIGHTS AND VARIABLE ANALYSIS:

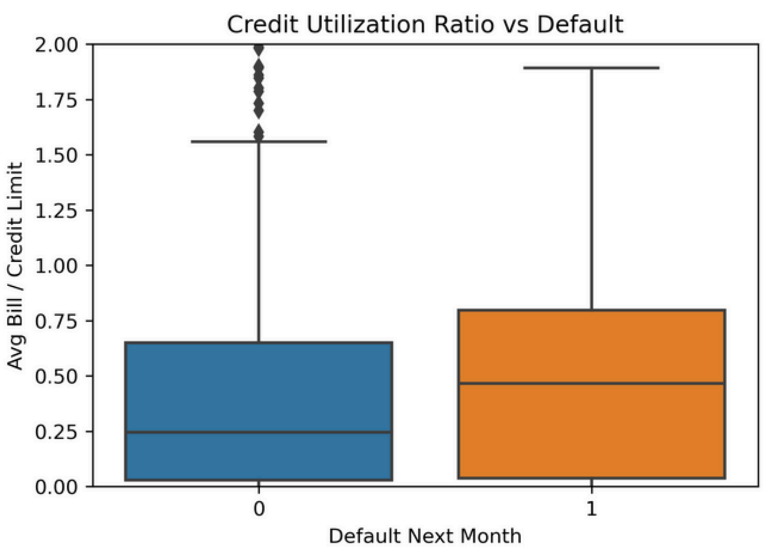
Delinquency Streak vs Default:

- Delinquency streak measures how many of the past 6 months a customer was overdue.
- Non-defaulters usually have 0–1 overdue months.
- Defaulters typically show 1–4 overdue months, with some reaching the full 6-month streak.
- The median streak for defaulters is much higher, and the overall distribution is wider.



Credit Utilization Ratio vs Default :

- The credit utilization ratio(bill / credit limit) is a key financial indicator that measures how much of a customer's available credit they are using.
- Defaulters tend to have higher credit utilization ratios, with a median around 0.5, and many cases exceeding 1.0 (over-limit usage)
- Non-defaulters have lower and more stable utilization, mostly between 0.2–0.6.

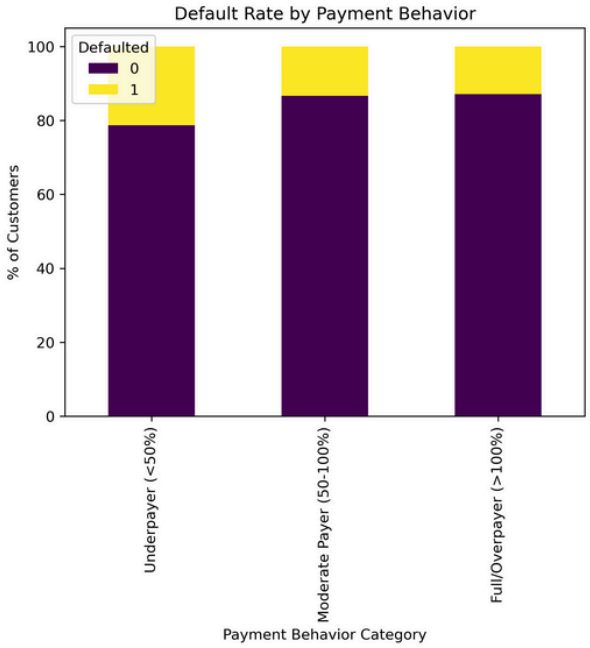


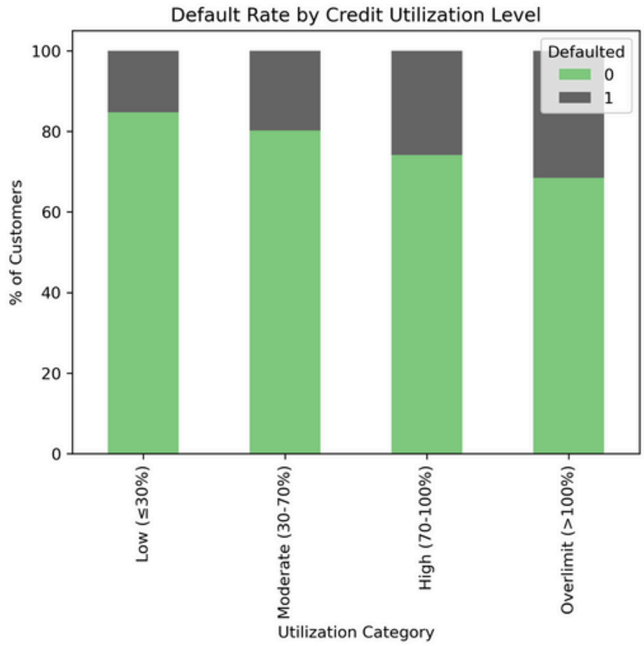
avg_utilization:

- On average, customers use 37% of their credit limit monthly.
- Some customers over-utilize credit (ratio >1), indicating high credit stress.
- Higher utilization slightly increases default risk.

total_pay:

- Lower payment volumes are weakly associated with higher default risk.





MODEL COMPARISON AND FINAL SELECTION JUSTIFICATION:

XGBoost selected as the final model because:

- It achieved the highest F1, F2, and ROC-AUC scores on the validation set.
- It generalizes well without overfitting as it has built-in regularization.
- It handled class imbalance effectively via SMOTE.
- Compatible with SHAP for explainability, helping interpret financial features like PAY_TO_BILL_ratio, delinquency_streak, and avg_utilization.

EVALUATION METHODOLOGY:

Primary Metric:

F2 Score:

- F2 Score captures as many actual defaulters as possible, which aligns with the bank's objective of preventive risk management.
  (The F2 Score gives more weight to recall than precision)
- False Negatives (FN) are the defaulters incorrectly labeled as non-defaulters.
- These are very costly for the bank because they lead to missed early warnings, poor risk provisioning and potential financial loss.

Secondary Metrics:

Recall:

- Measures how many actual defaulters the model correctly identified.
- Important to catch as many risky customers as possible.

Precision:

- Tells us how many flagged defaulters were actually defaulters.
- Helps reduce false alarms.

F1 Score:

- Balances precision and recall.

ROC-AUC:

- Shows the model's ability to separate defaulters from non-defaulters across all thresholds.
- A higher value means better overall classification quality.

METRICS RESULT ON TRAIN DATASET

- **Accuracy:** 0.89

89% of overall predictions were correct

- **F2 Score:** 0.8606

Prioritizes recall and ensures strong performance for risk tasks

- **Recall: (Class 1:Defaulters)**

High recall (84%) ensures most defaulters are captured, which is critical for credit risk.

- **F1 Score:** 0.888

Strong balance between precision and recall

- **ROC-AUC score:**0.947

Excellent ability to separate defaulters from non-defaulters

These results indicate the model is highly effective at identifying high-risk customers early, making it suitable for proactive credit risk management.

CLASSIFICATION CUTOFF SELECTION:

- In credit risk prediction, the cost of missing a defaulter (false negative) is much higher than wrongly flagging a non-defaulter (false positive).
- Therefore, objective was to maximize the model's ability to identify actual defaulters, even if it results in slightly more false alarms.

Threshold Optimization Process:

1. Generated prediction probabilities on the training set.
2. Evaluated performance at multiple thresholds ranging from 0.1 to 0.9.
3. Tracked the F2 Score across these thresholds since F2 prioritizes recall.
4. Identified the optimal threshold at 0.20, where the F2 Score peaked.

Threshold Tuning:

The best performance was observed at a threshold of 0.20

- Achieved high recall, minimizing missed defaulters
- Maintained a strong F2 Score, aligning with the bank's risk strategy.
- Balanced early risk detection with acceptable false positives.

This cutoff was used for final predictions on the validation set to ensure the model supports early intervention and proactive credit risk management.

BUSINESS IMPLICATIONS:

Early Warning System:

- Identifies likely defaulters before they miss payments.
- This enables timely interventions such as reminders, limit cuts, or loan restructuring.

Risk Based Customer Segmentation:

- Customers can be grouped based on their predicted default probability:

  Low risk customers → Eligible for higher limits or premium offers

  High risk customers → Monitored closely or placed under credit hold

Improved Credit Portfolio Health:

- Minimizes non-performing assets (NPAs) by reducing exposure to likely defaulters.
- Enhances regulatory compliance through better provisioning and risk governance.

Reduced Financial Loss:

- Automating the risk flagging process reduces the need for manual reviews.

Customer Experience Optimization:

- Responsible credit management avoids over-lending to risky profiles
- This reduces customer stress and improve satisfaction in the long term.

This model enables Bank A to take proactive, data-driven decisions that minimize credit losses, optimize customer engagement, and strengthen overall risk control.

SUMMARY OF FINDINGS AND KEY LEARNINGS

Findings:

Behavioral patterns are strong indicators of default:

Features such as payment status (PAY_0 to PAY_6), delinquency streaks, and low PAY_TO_BILL ratios are highly correlated with the likelihood of default.

High credit utilization signals financial stress:

Customers with a credit utilization ratio > 0.5 were significantly more likely to default, indicating that over-reliance on available credit is a key risk factor.

Defaulters have lower credit limits and weaker repayment histories:

Many defaulters were observed to have low LIMIT_BAL, high bill amounts relative to payments, and frequent past due records.

Class imbalance was substantial:
This was successfully handled using SMOTE, improving model performance on defaulters.

XGBoost outperformed all other models:
It achieved the best F2 Score (0.861), along with high recall (0.84) and ROC-AUC (0.947), making it ideal for credit risk tasks.

Threshold tuning (at 0.20) was crucial:
Lowering the classification threshold helped catch more actual defaulters, aligning model behavior with the bank's risk priorities.

Key Learnings:

Importance of Domain-Specific Features:
Financially meaningful features like credit_utilization_ratio and delinquency_streak significantly improved predictive power over raw variables alone.

Metric Choice Matters:
Prioritizing F2 Score over traditional metrics (like accuracy) ensured that the model aligned with real-world credit risk needs as it helps in catching more actual defaulters.

Class Imbalance Must Be Addressed:
Using SMOTE helped improve recall and stabilized performance across both classes in a heavily imbalanced dataset.

Threshold Tuning Is Critical:
Adjusting the classification threshold (to 0.20) rather than using the default 0.5 helped improve early detection of risky customers.

Iterative Evaluation Pays Off:
Systematic testing of multiple models (Logistic Regression, LightGBM, XGBoost) and configurations led to an evidence-based final model selection.

Explainability Builds Trust:
Tools like SHAP made the model more interpretable by highlighting which features influenced predictions, crucial for transparency in financial decisions.