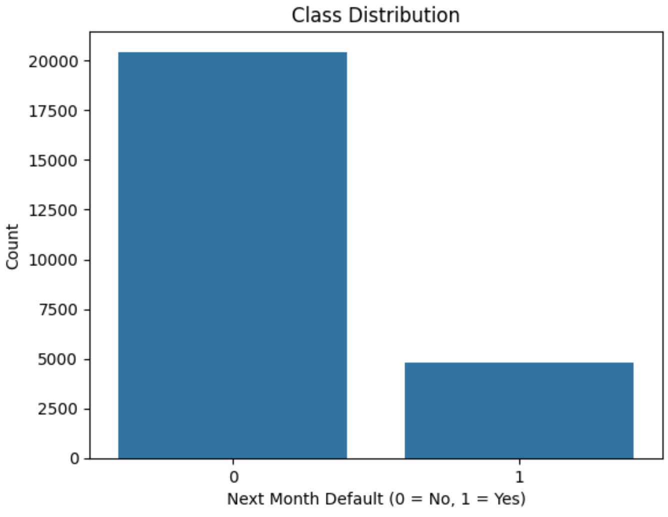
Name: Shreyansh Verma **Enrollment No:** 22118069

Credit Card Behaviour Score Prediction Using Classification and Risk-Based Techniques

1. Overview of Approach and Modeling Strategy

We aim to predict the probability of credit card default using customer behavioural data. The process includes:

- Cleaning and preprocessing the data.
- Exploratory Data Analysis (EDA) for identifying key patterns and relationships.
- Addressing class imbalance via SMOTE.
- Building and comparing multiple models: Logistic Regression, XGBoost, AdaBoost and LightGBM.
- Selected Logistic Regression based on F2 score and recall with a special emphasis on minimizing false negatives.
- **Class Imbalance:**



2. Feature Engineering: Added Columns & Their Significance

2.1 avg_bill_amt — Average Bill Amount

$$\text{avg\_bill\_amt} = \frac{1}{6}\sum_{i=1}^{6} \text{BILL\_AMT}_i$$

- **Purpose**: Captures a customer's **average financial obligation** per month.
- **Why it's important**:
  - Indicates the **typical monthly liability**.

- A higher average bill may suggest **overspending** or **higher lifestyle cost**, which could stress repayment capacity, especially if income is not proportionally high.

- **Finding**: Defaulters often had a **higher avg_bill_amt**.

2.2 credit_utilization — Credit Utilization Ratio

$$\text{credit\_utilization} = \frac{\text{avg\_bill\_amt}}{\text{LIMIT\_BAL} + 1}$$

- **Purpose**: Measures how much of the available credit is being used on average.
- **Why it's important**:
  - A high utilization rate (>50%) is a **known red flag in credit scoring**.
  - It reflects **financial stress**, **debt dependency**, or **overspending** behaviour.
- **Finding**: Defaults increased sharply with higher utilization.

2.3 delinquency_streak — Longest Consecutive Delays

- **How it's calculated**: Based on the maximum streak of positive values in PAY_0 to PAY_6.
- **Purpose**: Measures the **consistency of delayed payments**.
- **Why it's important**:
  - A high streak means the customer **regularly defaulted** over multiple months.
  - This is a **strong predictor of future default** due to sustained poor financial discipline.

2.4 min_payment_count — Count of Months with Minimum Payments

- **How it's calculated**: Count of months where payment >= minimum due (or consistently small but non-zero payments).
- **Purpose**: Measures **repayment compliance**, even if only minimum dues are paid.
- **Why it's important**:
  - Shows **intent to pay** or effort to avoid default.
  - A higher count of such payments typically correlates with **lower risk**, even if full payment is not made.

2.5 repayment_variability — Standard Deviation of PAY_x Series

$$\text{repayment\_variability} = \text{std}(PAY\_0, PAY\_2, \ldots, PAY\_6)$$

- **Purpose**: Measures the **stability of repayment behaviour**.
- **Why it's important**:
  - Higher standard deviation = **inconsistent repayment**.
  - Irregular repayment schedules (sometimes paying early, sometimes not at all) reflect **financial instability**.
- **Finding**: Higher variability was associated with more defaults.

2.6 overpayment_count — Number of Months Where Payment > Bill

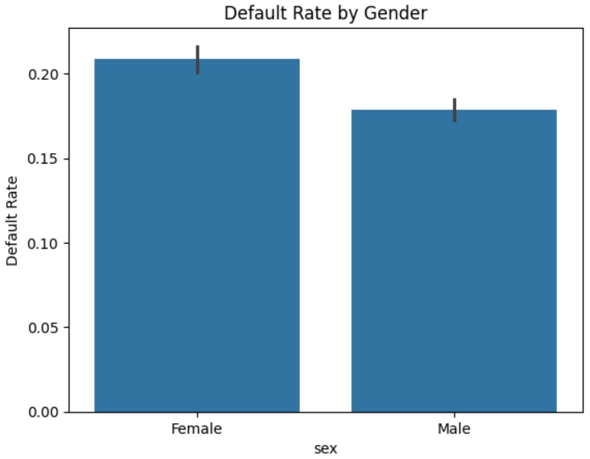
$$\text{overpayment\_count} = \text{count of months where } PAY\_AMT_i > BILL\_AMT_i$$

- **Purpose**: Measures **proactive payment behaviour**.
- **Why it's important**:
  - Overpayments show **strong financial discipline** and **buffer-building**.
  - Customers who frequently overpay are **less likely to default**.
- **Finding**: Default rate declined as overpayment count increased.

3. EDA Findings and Visualizations

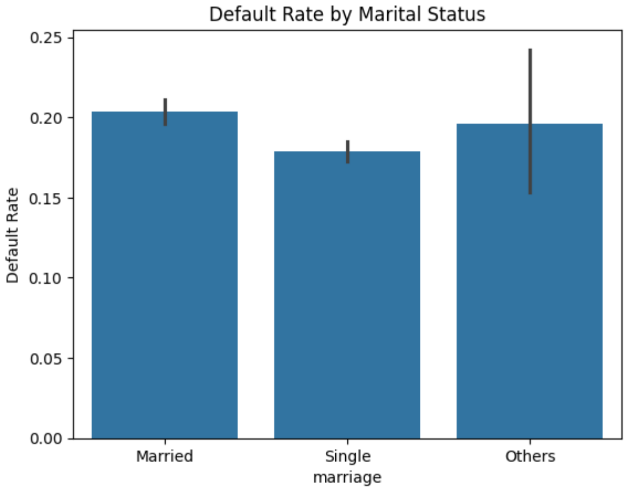
3.1 Gender vs Default

- **Observation**: Males had a slightly higher default rate compared to females.
- **Reasoning**: This could reflect behavioural patterns, risk appetites, or financial planning differences between genders.



3.2 Marital Status vs Default

- **Observation**: Single individuals had the highest default rate, followed by others, with married people having the lowest.
- **Interpretation**: Single people may face higher financial instability or lack of dual income buffers, increasing credit risk.

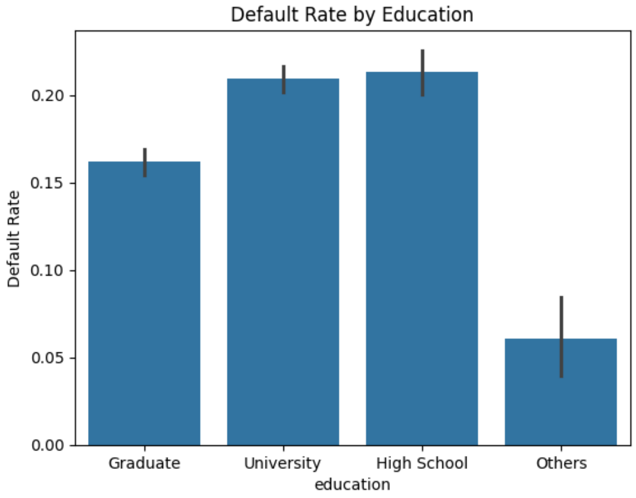


3.3 Education vs Default

- **Findings**:

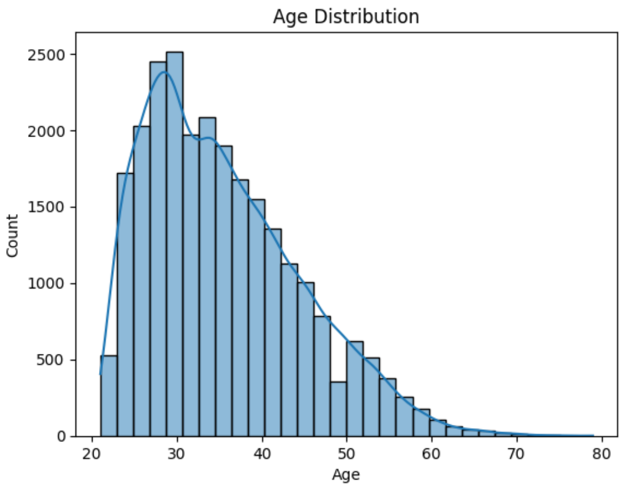
  - "High School" and "Others" categories showed a higher rate of default.
  - "Graduate" and "University" groups were more financially stable.

- **Implication**: Lower education correlates with potentially lower income and poor financial awareness, driving defaults.



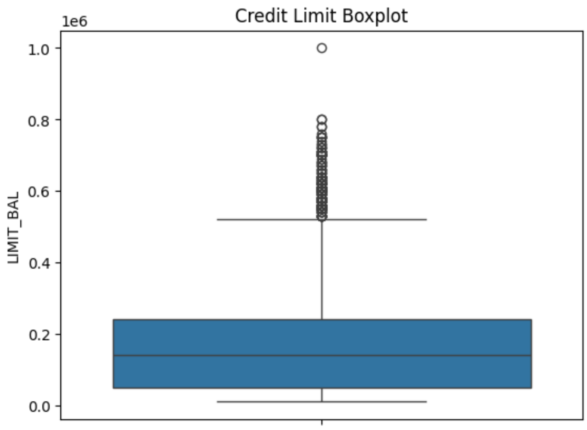
3.4 Age Distribution

- **Insight**: Most customers were between 25–45 years old.
- **Trend**: The age group 30–35 was dominant; no clear age bias for default was seen in the distribution, but younger customers might slightly skew toward default.



3.5 Credit Limit Boxplot

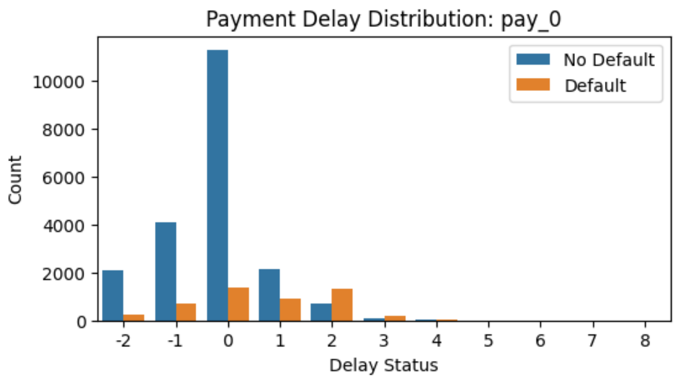
- **Summary**: Wide range of credit limits, with many outliers at the upper end.
- **Trend**: While defaults exist across all credit limits, those with low limits tended to default more, implying financial stress or lower creditworthiness.

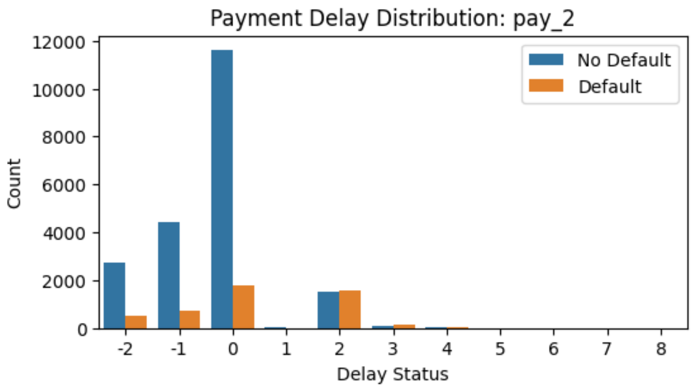


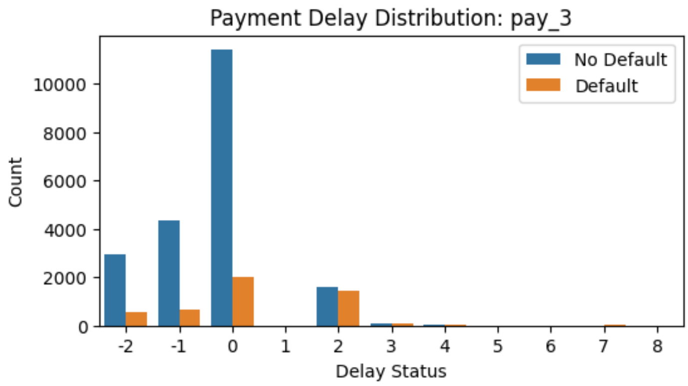
3.6 Payment Delay Status Visualizations (pay_0 to pay_6)

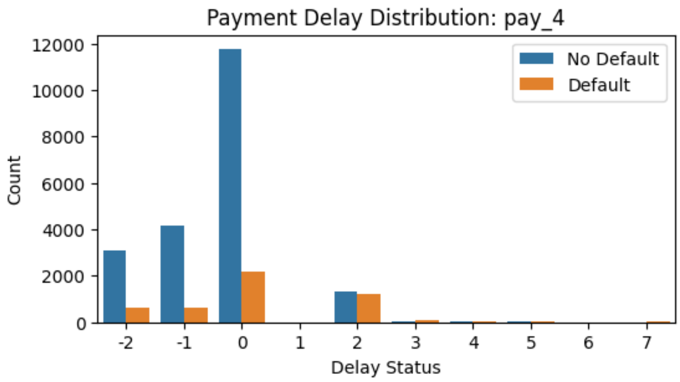
- **Observation**: As delay codes increase (e.g., -1 = early, 0 = on time, 1+ = late), default probability increases.
  - ➢ PAY_0 and PAY_2 have the highest correlation with default.
  - ➢ Higher delay values (≥1) strongly correlate with default.
  - ➢ Users consistently delaying payments (even by 1 month) are higher risk.

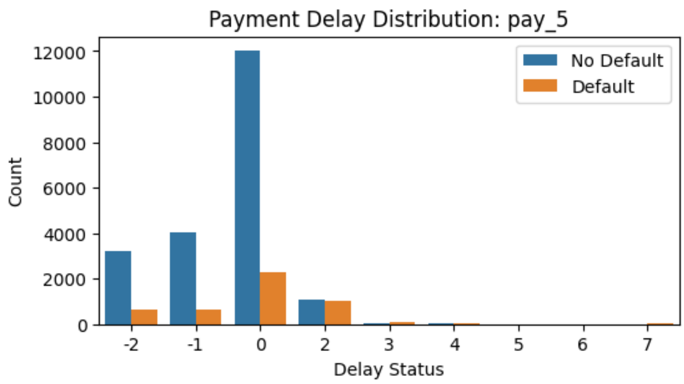
- **Financial Insight:** Higher delay values (≥1) are strongly associated with default. PAY_0 (most recent month) is the most influential.
- **Example**: A payment delay of 3+ months consistently showed very high default frequencies.
- **Conclusion**: Repayment history (i.e., timeliness) is a **strong predictor** of future default.

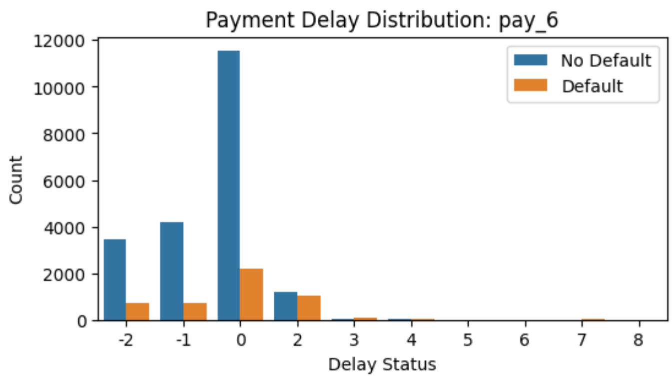






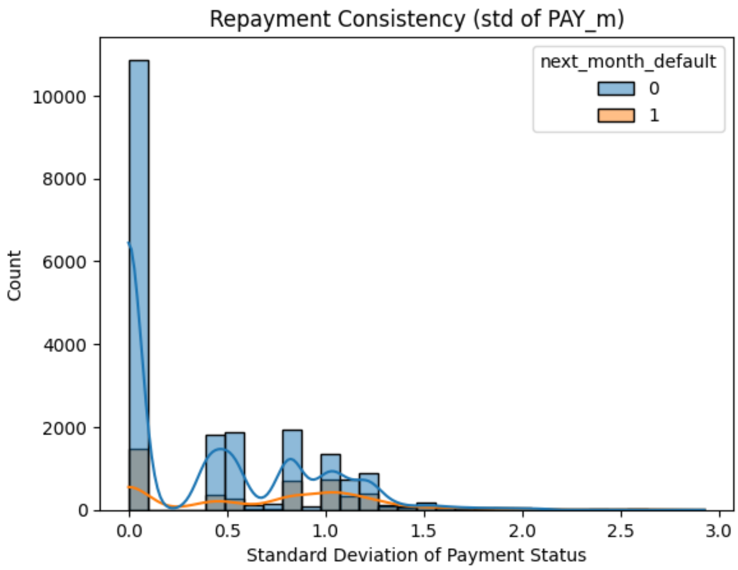






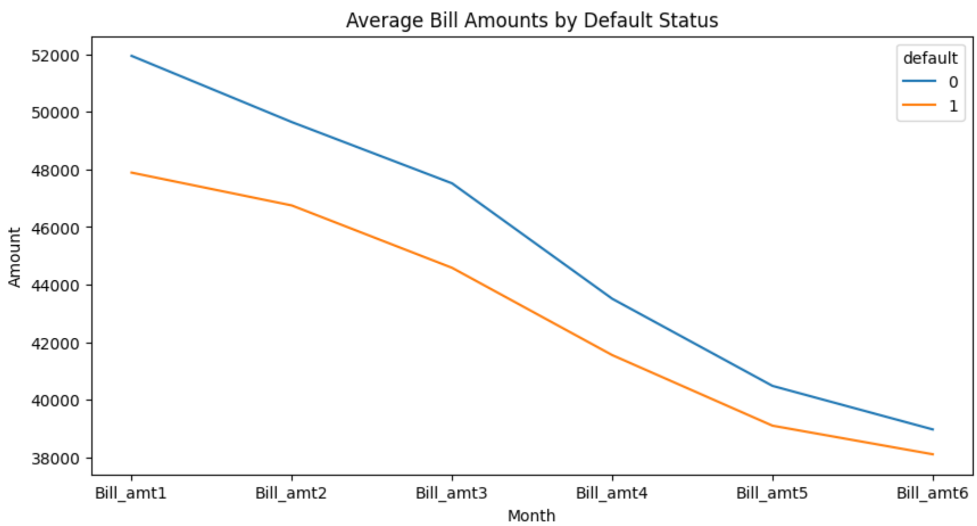
3.7 Repayment Consistency

- **Metric Used**: Standard deviation of payment delay (repayment_variability).
- **Insight**: Higher variability → irregular repayment behaviour → more likely to default.
- **Outcome**: Defaults were concentrated among customers with **high volatility** in repayment, showing behavioural instability.



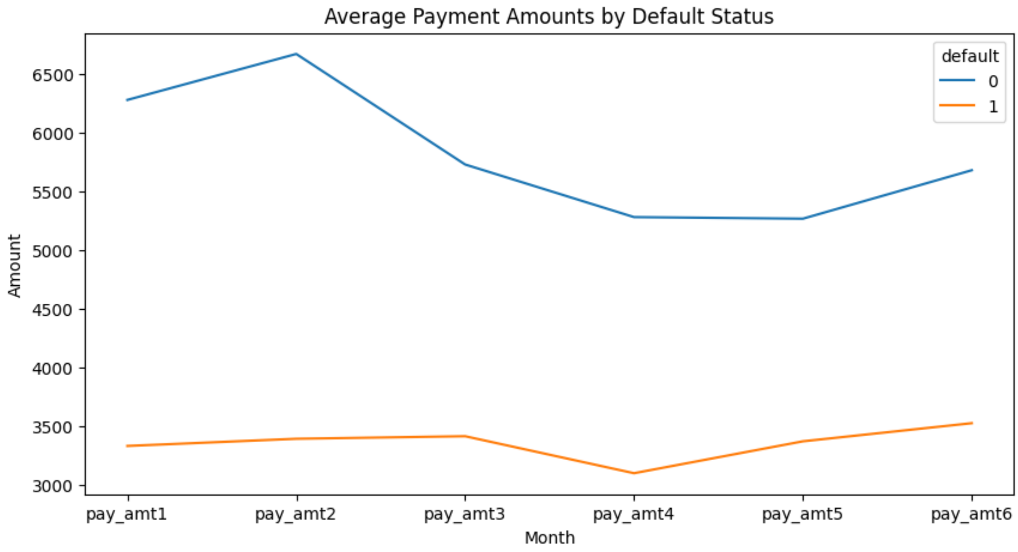
3.8 Average Bill Amounts vs Default Status

- **Trend**: Defaulters often had **consistently higher average monthly bills** across 6 months.
- **Interpretation**: This may indicate overspending or poor budget control.
- **Implication**: Elevated financial obligations without matching income/payment capacity.



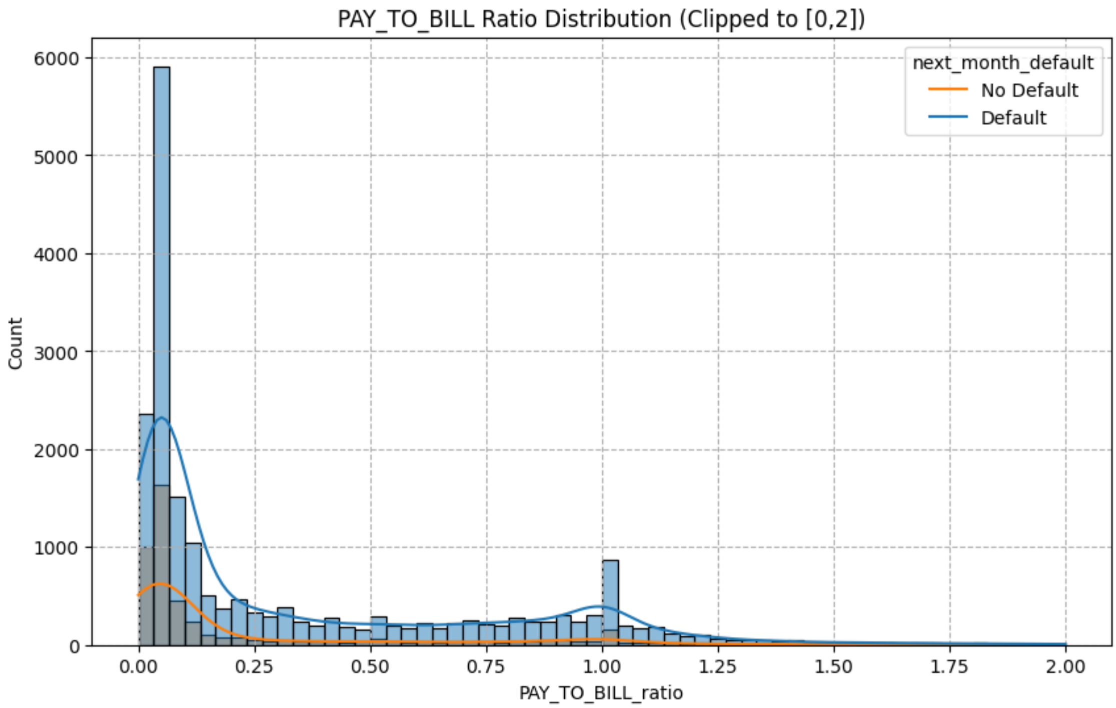
3.9 Average Payment Amount vs Default Status

- **Insight**: Non-defaulters consistently paid **higher average amounts** than defaulters.
- **Trend**: Defaulting users' payment curves often flattened, suggesting missed or minimal payments.



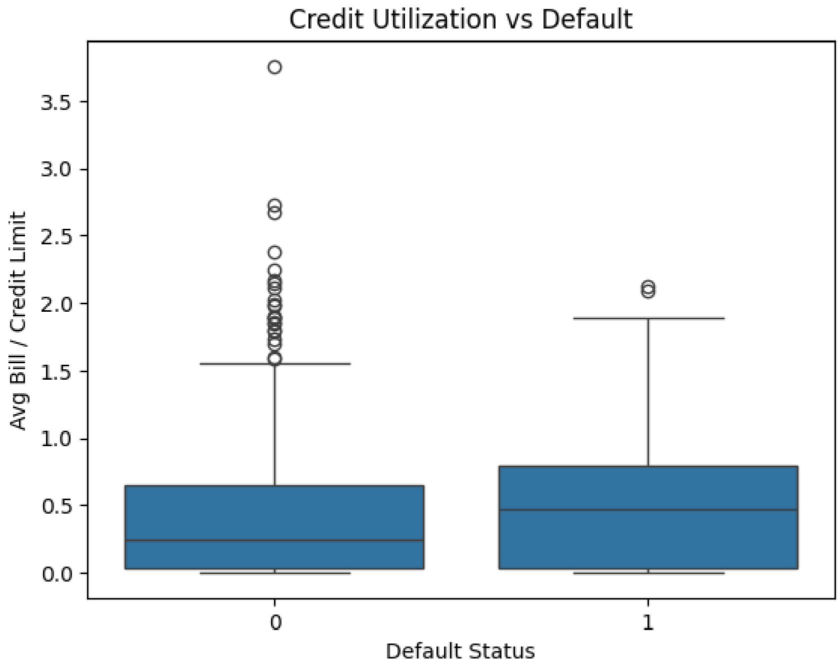
3.10 Pay to Bill Ratio Distribution

- **Definition**: Ratio of monthly payment to the monthly billed amount.
- **Insight**: Defaulters had **lower ratios**, frequently < 0.5, indicating they didn't cover even half their dues.
- **Distribution**: Long left tail for defaulters and more symmetrical for non-defaulters.



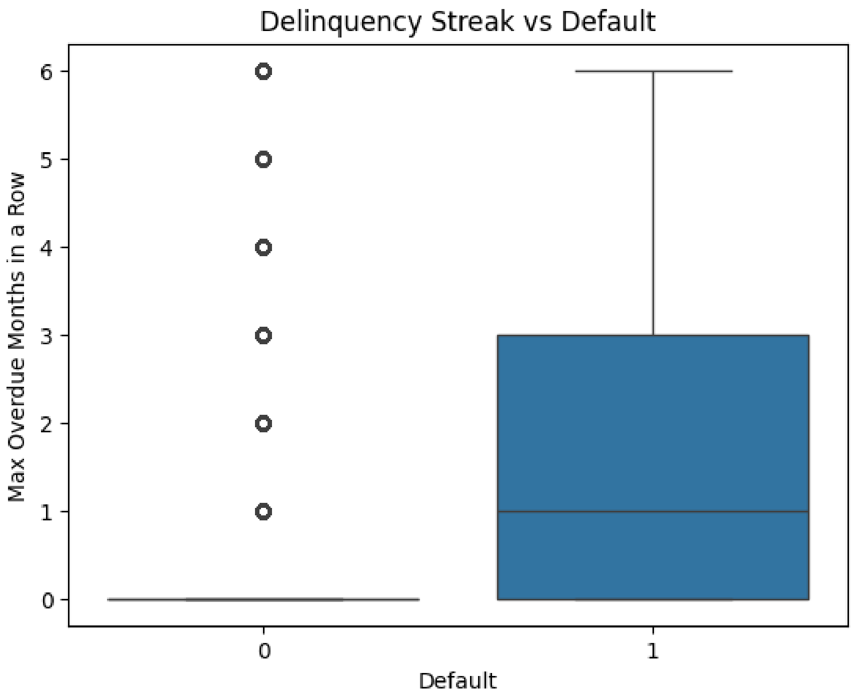
3.11 Credit Utilization vs Default

- **Definition**: avg_bill_amt / credit_limit
- **Finding**: Defaults increased sharply when utilization exceeded ~50%.
- **Interpretation**: Higher utilization ratio → financially stretched → higher probability of default.



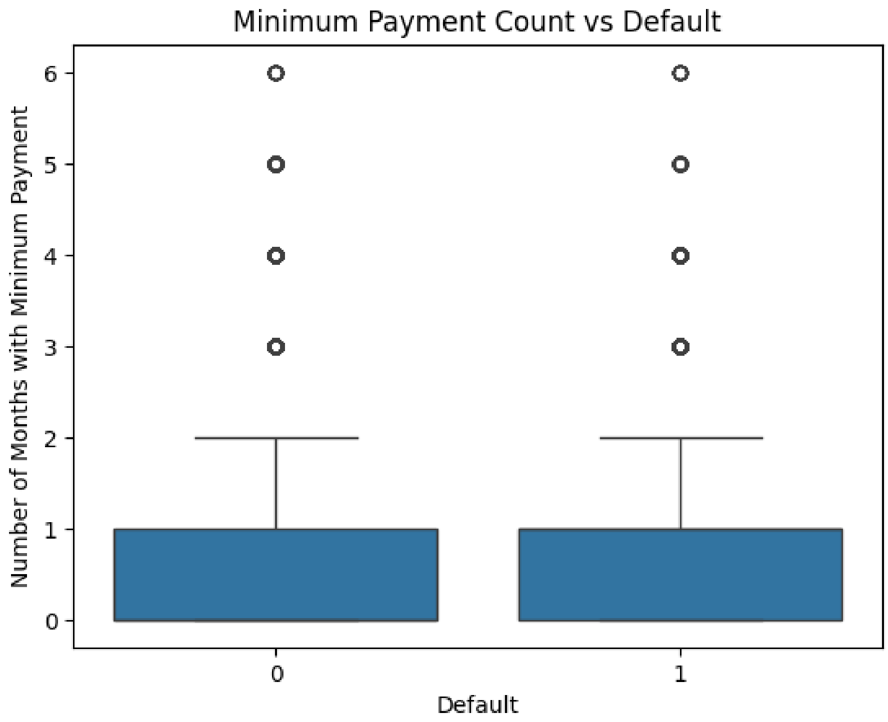
3.12 Delinquency Streak vs Default

- **Variable**: Longest stretch of delayed payments.
- **Observation**: Users with a streak of 3+ months of delayed payments had much higher default rates.
- **Conclusion**: Chronic late payers are a strong risk segment.



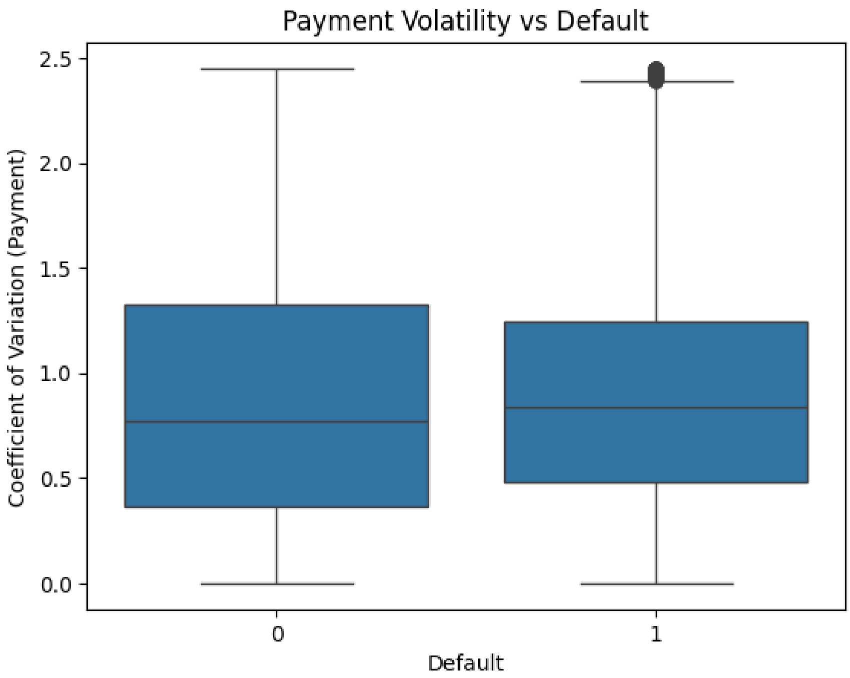
3.13 Minimum Payment Count vs Default

- **Metric**: Number of months minimum payment was made.
- **Trend**: Non-defaulters had a **higher count** of months with minimum payments.
- **Implication**: Regular small payments help avoid default, even if full dues aren't paid.



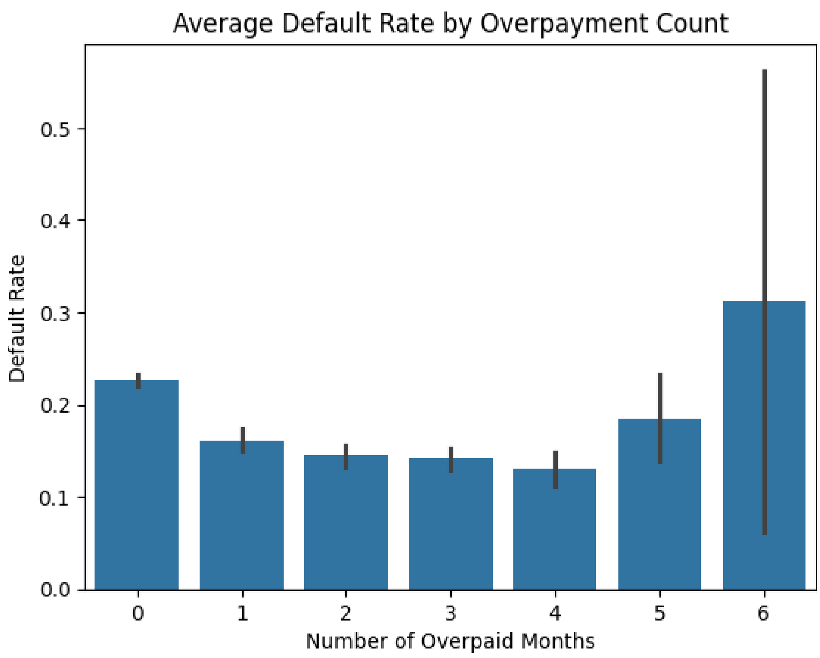
3.14 Payment Volatility vs Default

- **Definition**: Standard deviation in payment amounts over time.
- **Finding**: Higher volatility → increased default risk.
- **Reasoning**: Irregular payments reflect unpredictable finances or income variability.



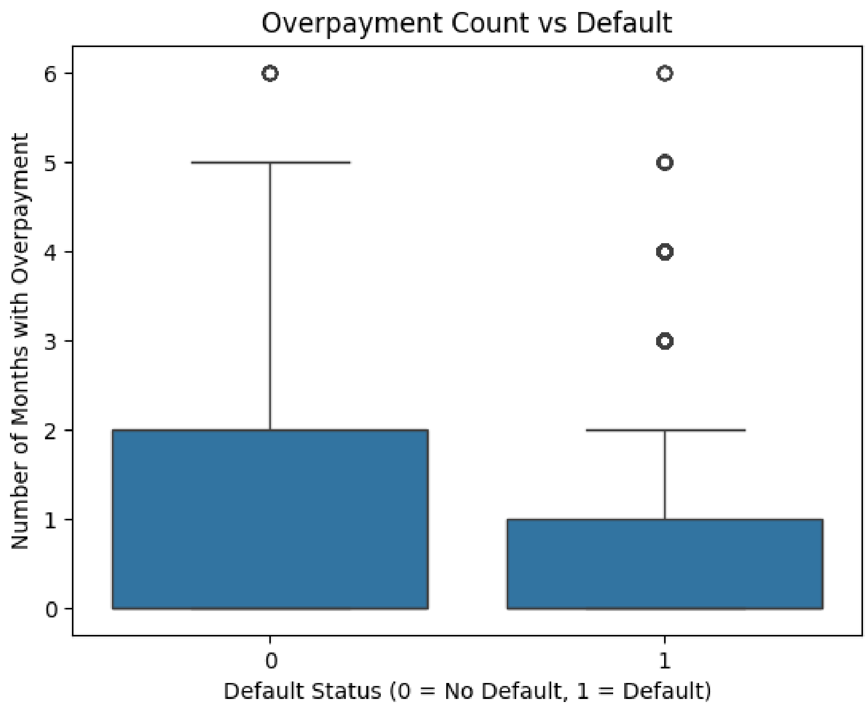
3.15 Average Default Rate by Overpayment Count

- **Variable**: Number of times the customer paid more than the billed amount.
- **Trend**: Default rate **dropped sharply** as overpayment count increased.
- **Interpretation**: Overpayment signals strong financial discipline.

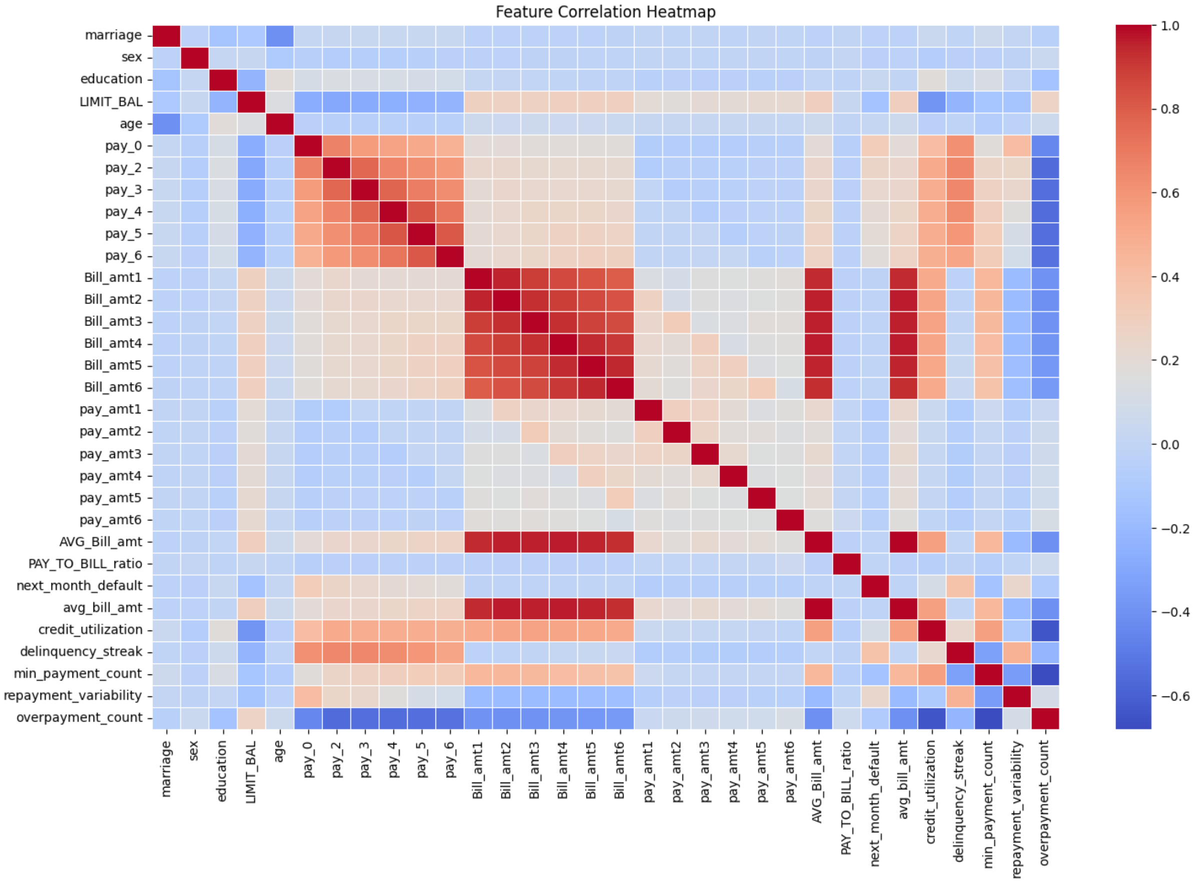


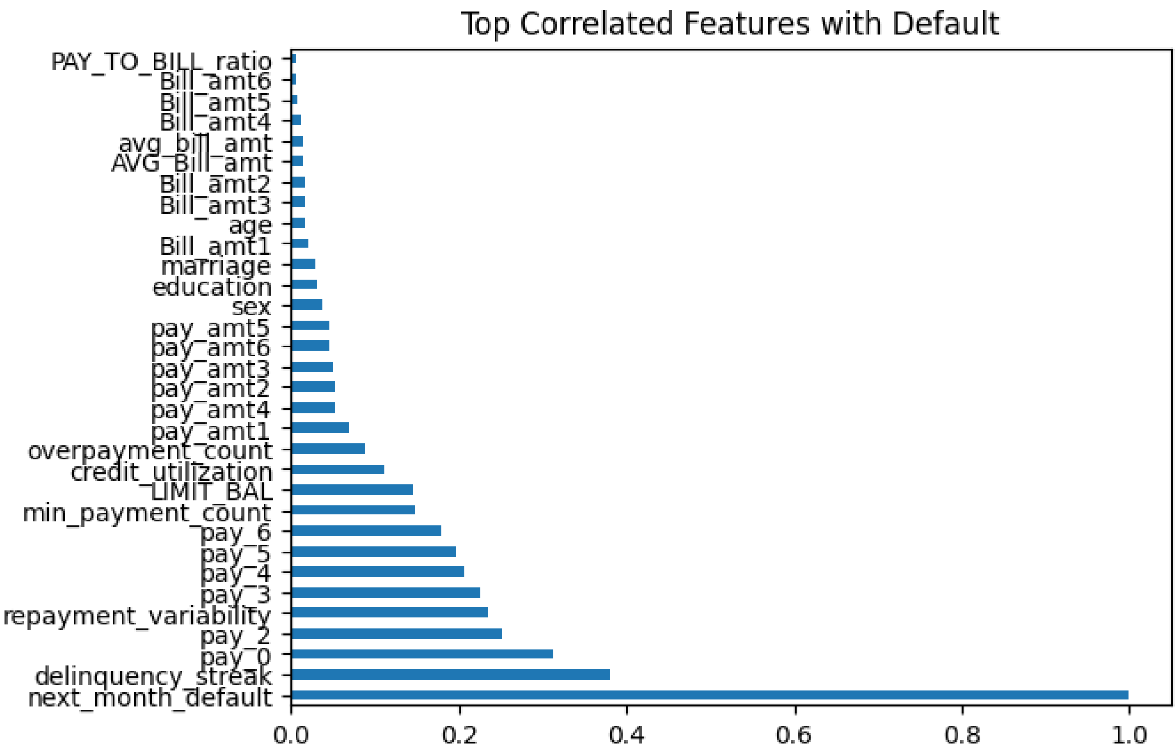
3.16 Overpayment Count vs Default

- **Observation**: Most defaulters had **zero or very few** overpayments.
- **Conclusion**: Overpayment is a leading indicator of lower credit risk.



3.17 CORRELATION HEATMAP:





4. Model Selection:

During model experimentation, multiple classification models were trained and evaluated — including Decision Trees, Random Forest, and possibly others — on both the full dataset and a reduced top 10 feature set derived from SHAP values. While these alternative models can capture non-linear patterns, the final choice was **Logistic Regression** trained on the **full feature set** because:

4.1 Logistic Regression Offers Better Generalization with Interpretability:

- Logistic regression provides a linear decision boundary and probabilistic outputs, making it highly interpretable.

- With well-engineered features (like credit utilization, repayment variability), logistic regression can achieve high accuracy without overfitting — especially valuable when the goal is to detect general trends in credit default behaviour.

- Unlike tree-based models which can be prone to overfitting, logistic regression remains stable and generalizable on unseen data, especially when the number of features is relatively high.

4.2 Using Full Feature Set Helped Retain Predictive Patterns

- Two training strategies were tested:
  - Top 10 features (based on SHAP importance)
  - All available features + engineered features
- The full dataset consistently outperformed the reduced version in terms of F1 score, recall, and F2 score.
- Conclusion: Limiting the model to only 10 features caused it to miss subtle but meaningful patterns captured by other variables. Since model complexity wasn't a constraint, using the full set gave the most accurate and robust results.

4.3 Focus on Recall over Precision (Why Threshold = 0.4)

- In credit risk modeling, especially for default prediction, recall is often more critical than precision:

  - Recall ensures we catch most defaulters (i.e., reduce false negatives).

  - Precision ensures we don't wrongly label too many good customers — but this is less costly than missing an actual defaulter.

- Therefore, the classification threshold was adjusted from 0.5 to 0.4:

  - This increased sensitivity of the model, enabling it to identify more defaulters at the cost of a slight increase in false positives.

  - A defaulting customer has a much higher cost implication than wrongly flagging a non-defaulter.

4.4 Why Not Use More Complex Models Like Random Forest / XGBoost?

- Tree-based ensemble models (e.g., RF, XGBoost) often deliver higher raw accuracy, but:

- They are harder to interpret, especially in regulated environments (e.g., finance).
- They can overfit, especially when the feature set is noisy or correlated.
- Model explainability is often a key factor in credit risk applications, as decisions may need to be explained to stakeholders or regulators.

- Logistic Regression combined with SHAP analysis gives a good balance of interpretability and performance**.**

5. Performance Metrics Considered

- The model was evaluated based on:
  - **F1 Score**: Balanced metric combining precision and recall.
  - **Recall**: Prioritized metric for identifying risky customers.
  - **F2 Score**: Heavily recall-focused, further supporting our objective.
  - **Accuracy**: Monitored but not prioritized due to class imbalance.

```
==== Logistic Regression Evaluation ====
Accuracy: 0.7576237623762376
F1 Score: 0.49169435215946844
F2 Score: 0.5613502749857766
Recall: 0.6198952879581152
Precision: 0.4074328974535444
ROC AUC: 0.7676292759015272
```

```
==== XGBoost Evaluation ====
Accuracy: 0.8277227722772277
F1 Score: 0.45962732919254656
F2 Score: 0.4134078212290503
Recall: 0.387434554973822
Precision: 0.5648854961832062
ROC AUC: 0.7601593054996197
```

```
==== LightGBM Evaluation ====
Accuracy: 0.837029702970297
F1 Score: 0.4714193962748876
F2 Score: 0.4149706015377657
Recall: 0.3842931937172775
Precision: 0.6096345514950167
ROC AUC: 0.782402239993863
```

```
AdaBoost (Threshold=0.4)
              precision    recall  f1-score   support

           0       0.96      0.10      0.19      4095
           1       0.20      0.98      0.34       955

    accuracy                           0.27      5050
   macro avg       0.58      0.54      0.26      5050
weighted avg       0.81      0.27      0.22      5050

F2 Score: 0.5556874851579198
```

Focusing on recall in default prediction tasks is essential because of the real-world costs and consequences associated with failing to identify defaulters.

★ Missing a Defaulter is Costly

- False Negatives (defaulters predicted as non-defaulters) are financially risky:
  - The bank or credit issuer may approve a loan or credit line assuming the customer is low-risk.
  - If that customer defaults, it results in direct financial loss, recovery costs, and opportunity cost.
- In contrast, a False Positive (predicting a good customer as a defaulter):
  - May result in denying a good customer credit — this is bad for business, but not as financially damaging as a real default.

★ Real-World Trade-Off: Recall vs Precision

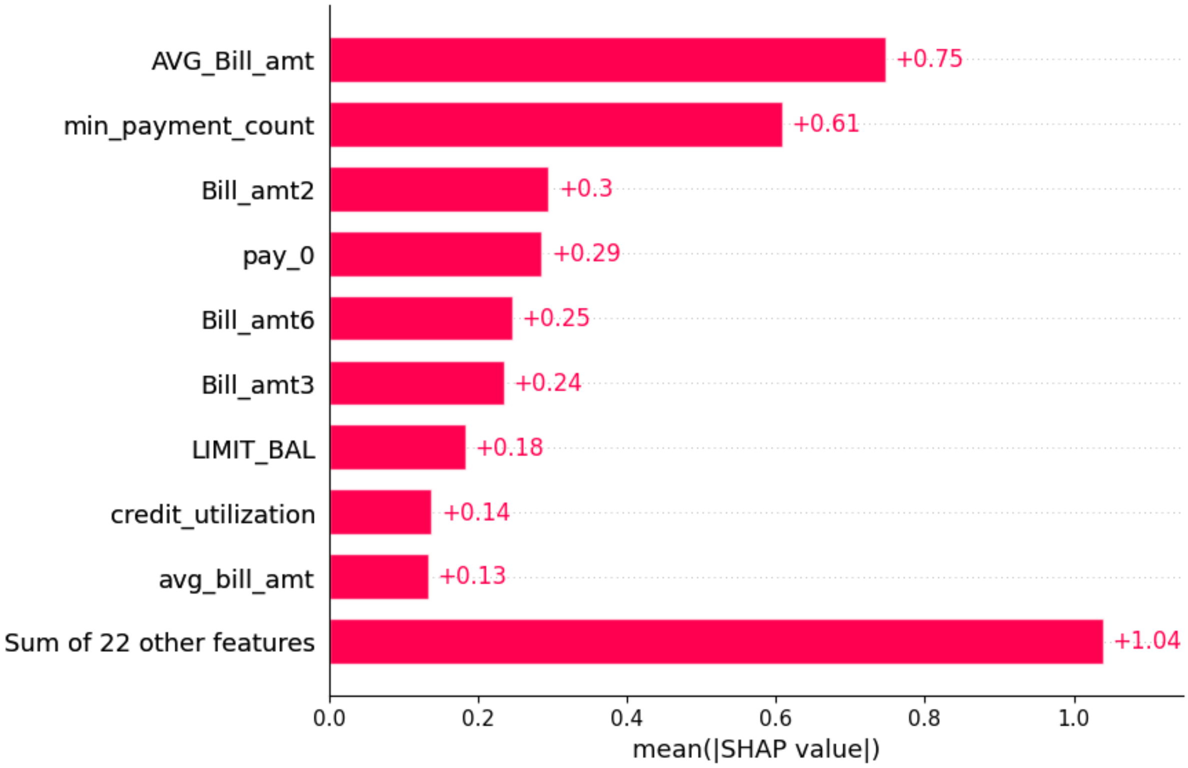
- Increasing recall might reduce precision, i.e., more false alarms (flagging non-defaulters as risky).
- But this trade-off is acceptable in credit risk, because:

- ➢ Losing a few good customers is manageable.
- ➢ Missing a true defaulter could mean losing thousands to lakhs of rupees or more per customer.

- This trade-off is reflected in the business cost function, where false negatives cost more than false positives.

6. SHAP Analysis:

- The model is heavily reliant on financial behaviour-based features, rather than purely demographic variables.
- SHAP confirms that recent bill amounts**,** repayment patterns, and credit usage behaviour are far more predictive than static variables like age or gender.
- The feature engineering efforts (e.g., AVG_Bill_amt, min_payment_count, credit_utilization) significantly enhance model interpretability and predictive power.



- This SHAP plot validates the effectiveness of engineered features and aligns well with domain knowledge of credit risk:
  - ➢ High bills + minimum payments → likely defaulter
  - ➢ Timely payments + low utilization → low risk

7. Summary of Findings

After a comprehensive end-to-end analysis of credit card default prediction using a real-world dataset, the following key insights and learnings were drawn:

7.1 Data-Driven Insights into Default Behaviour

- **Payment History is Critical**: Features like *PAY_0, PAY_2*, and *PAY_3* (indicating recent repayment delays) emerged as strong predictors of default. Even a single late payment notably increased the likelihood of default.
- **Average Bill Amount Matters**: Users with consistently high billing amounts (*AVG_Bill_amt*) were significantly more prone to default, especially when not matched by proportionate repayments.
- **Minimum Payment Behaviour Signals Risk**: The engineered feature *min_payment_count* — counting months with only minimum payments — showed a strong positive correlation with default risk.
- **Credit Utilization is a Leading Indicator**: High *credit_utilization* (average bill / credit limit) was linked with a sharp increase in default probability, aligning well with domain knowledge.

7.2 Feature Engineering Made a Difference

- Custom features like *repayment_variability, delinquency_streak*, and *overpayment_count* added depth to customer behaviour modeling:

  - ➢ High repayment volatility indicated financial instability.
  - ➢ Frequent overpayments correlated with lower risk, showing proactive repayment tendencies.
  - ➢ Long delinquency streaks strongly flagged chronic defaulters.

These features boosted both predictive accuracy and model interpretability, proving the value of thoughtful feature engineering in credit risk tasks.

7.3 SHAP Explained the Model Behaviour Transparently

- SHAP analysis confirmed that the model's predictions were based on logical and intuitive financial patterns:

  - ➢ High billing with inconsistent payments pushed predictions toward default.
  - ➢ Stable and proactive financial behaviour pushed predictions away from default.

- The model prioritized repayment history and spending behaviour over static demographics like gender or education, which had minimal SHAP impact.

7.4 Model Choice and Thresholding Strategy

- After comparing multiple models (Logistic Regression, Random Forest, XGBoost, etc.), Logistic Regression on the full feature set was chosen:

  - ➢ Provided high interpretability, regulatory alignment, and solid recall performance.
  - ➢ Despite higher complexity, using the full dataset ensured no important patterns were missed.

- A probability threshold of 0.4 was chosen instead of 0.5 to increase recall, prioritizing the identification of defaulters even at the expense of a few false positives. This is vital in financial domains where missing a defaulter is costlier than flagging a non-defaulter.

7.5. Business Implications

- The final model can help credit issuers proactively flag risky customers, reducing potential loss from unpaid credit.
- Provides interpretable reasoning behind customer risk, which is crucial for compliance and customer communications.

- Can be integrated into real-time risk scoring systems to aid decisions on:

    - Credit limit increases
    - Loan approvals
    - Collections strategy
    - Offering hardship programs for at-risk customers