Classification Report:

	precision	recall	f1-score	support
0	0.68	0.54	0.60	1201
1	0.62	0.74	0.68	1201
accuracy			0.64	2402
macro avg	0.65	0.64	0.64	2402
weighted avg	0.65	0.64	0.64	2402

ROC AUC: 0.6815

> 6004 7(a) eligible out of 100000. Undersampled the data to 6004 eligible and 6004 ineligible.

Classification Report:

	precision	recall	f1-score	support
0	0.92	0.41	0.57	119132
1	0.62	0.96	0.76	119131
accuracy			0.69	238263
macro avg	0.77	0.69	0.66	238263
weighted avg	0.77	0.69	0.66	238263

ROC AUC: 0.7013

595,656 7(a) eligible out of 1,00,00,000. Undersampled the data to 595,656 eligible and 595,656 ineligible.

Things to be done:
Solving sampling issues, Cross-validation, Hyperparameter tuning

key assumptions for each column:
Business Structure & Ownership

1. Business Structure
   - Weighted distribution: Sole Proprietorship (25%), LLC (44.2%), S-Corp (25%), C-Corp (0.8%), Partnership (2%), Franchise (1.5%), Nonprofit (0.5%), Joint Venture (0.5%)
   - Nonprofit status automatically sets "For Profit" = False
2. Ownership Distribution
   - 55% single owner, 35% two owners, 7% three owners, 2% four owners, 1% five owners
   - Minimum 20% ownership per guarantor (SBA rule)
   - Ownership percentages sum to 100%

Credit & Financial Metrics

3. Credit Scores
   - Personal Credit (FICO 8): 300-850, normal distribution around 670
   - Business Credit (SBSS): 0-300, focused on 100-200 range
4. Financial History
   - Revenue: $150K-$5M (rounded to nearest $1K)
   - DSCR: 0.4-2.0 range, normal distribution around 1.25
   - Debt Calculation: Business Debt = NOI / DSCR
   - Net Profit Margin: Derived from (NOI/Revenue)×100
5. Temporal Features
   3-year history maintained for:
   - Revenue
   - Debt
   - NOI
   - DSCR

Loan Characteristics

6. Loan Amount
    - Microloans: $25-150K (10%)
    - Express: $150-500K (25%)
    - Standard: $500K-1M (40%)
    - Large: $1-5M (20%)
    - Edge Cases: Exact $25K/$500K/$5M (5%)
7. Collateral
    - Required for loans >$25K (with 5% exceptions)
    - Auto-calculated from loan amount

Geographic & Compliance

8. Location
    - 150+ US cities (no international locations)
    - Random selection with uniform distribution
9. NAICS Codes
    - Pre-filtered to exclude ineligible industries
    - 200+ eligible codes across sectors

Experience & Operations

10. Experience Metrics
    - Industry Experience: 0-20 years (uniform)
    - Managerial Experience: 0-20 years (uniform)
    - Years in Business: 0-15 years (uniform)

Loan Purpose Flags

11. Purpose Indicators
    - 14 binary flags with acquisition request logic:
        - Acquisition requests exclude other purposes
        - 5% random exceptions to collateral rules
    - Common purposes: Working Capital, Equipment Purchase, Expansion

Data Validation

12. Key Checks

- Ownership percentages sum to 100%
- Debt ≤ Revenue/DSCR
- NOI = DSCR × Business Debt
- Profit margins capped at 100%
- Loan amounts within SBA limits ($25K-$5M)

13. Exception Handling
- 5% boundary cases for loan amounts
- 5% collateral exceptions
- Realistic city/NAICS code filtering

Key Considerations

1. Regulatory compliance with SBA 20% ownership rule
2. Interdependent financial metrics for realism
3. Temporal consistency in 3-year financial histories
4. Weighted distributions reflecting real-world prevalence
5. Edge cases for robust model testing
6. Automatic flagging of nonprofit status
7. Program-specific eligibility enforcement