credit score

purpose :developing a credit score model to assess credit risk and support informed loan approval decisions

result : Provided insights into the impact of different cutoff scores on acceptance rates, bad rates, and profitability, enabling informed decision-making for lenders.

methodology : model

use OptBinning library in python to do all this

- bin each independent variable
- calculate each bin's information value (IV)
- calculate the log of (% of non-events / % of events) as the woe value
- replace each observation of the independent variable with the woe value
- perform WOE logistic regression on the dependent variable against the new woe dataset that has a monotonic trend in woe, and select variables based on IV
- evaluate model and create credit score

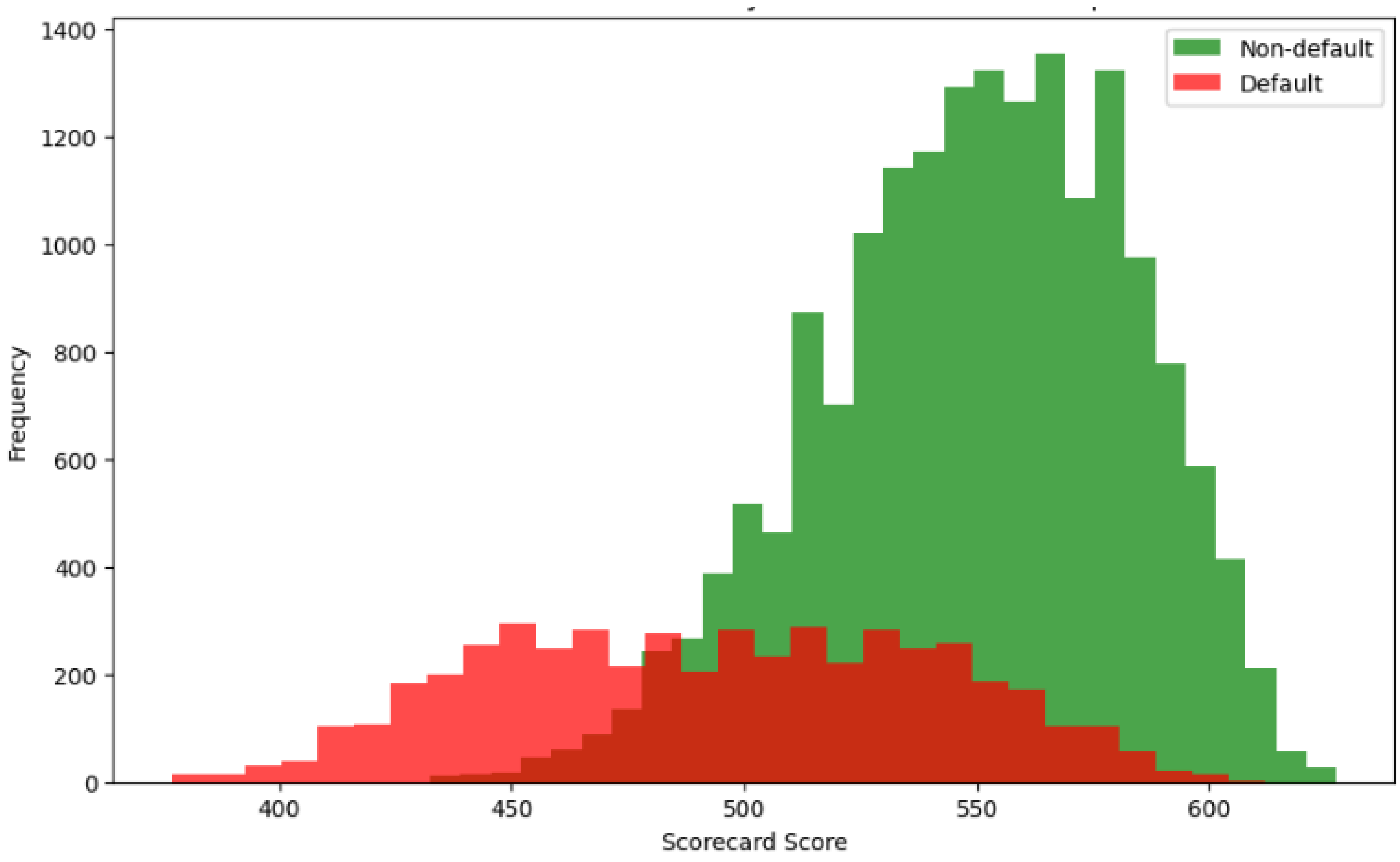
result of model

- Variable: person_income, Coefficient: -0.6266
- Variable: person_home_ownership, Coefficient: -0.8845
- Variable: person_emp_length, Coefficient: -0.2723
- Variable: loan_intent, Coefficient: -1.2654
- Variable: loan_percent_income, Coefficient: -0.9827
- Variable: cb_person_default_on_file, Coefficient: -1.1686
- auc = 0.82
- gini = 0.65

example of credit score from woe logistic regression

	Variable	Bin	Points
0	person_income	(-inf, 23002.00)	56.15
1	person_income	[23002.00, 34986.00)	72.57
2	person_income	[34986.00, 39937.50)	83.40
3	person_income	[39937.50, 49986.00)	87.70
4	person_income	[49986.00, 59846.50)	88.34
5	person_income	[59846.50, 69371.50)	94.92
6	person_income	[69371.50, 79942.50)	95.27
7	person_income	[79942.50, 108810.00)	105.99
8	person_income	[108810.00, inf)	107.44

By assigning higher score points to higher income bins, the scoring system captures the intuitive relationship between income and risk: as income increases, the risk of default or the event of interest decreases. This monotonic trend in the score points allows the income variable to be used effectively in credit scoring or risk modeling, where higher scores indicate lower risk.



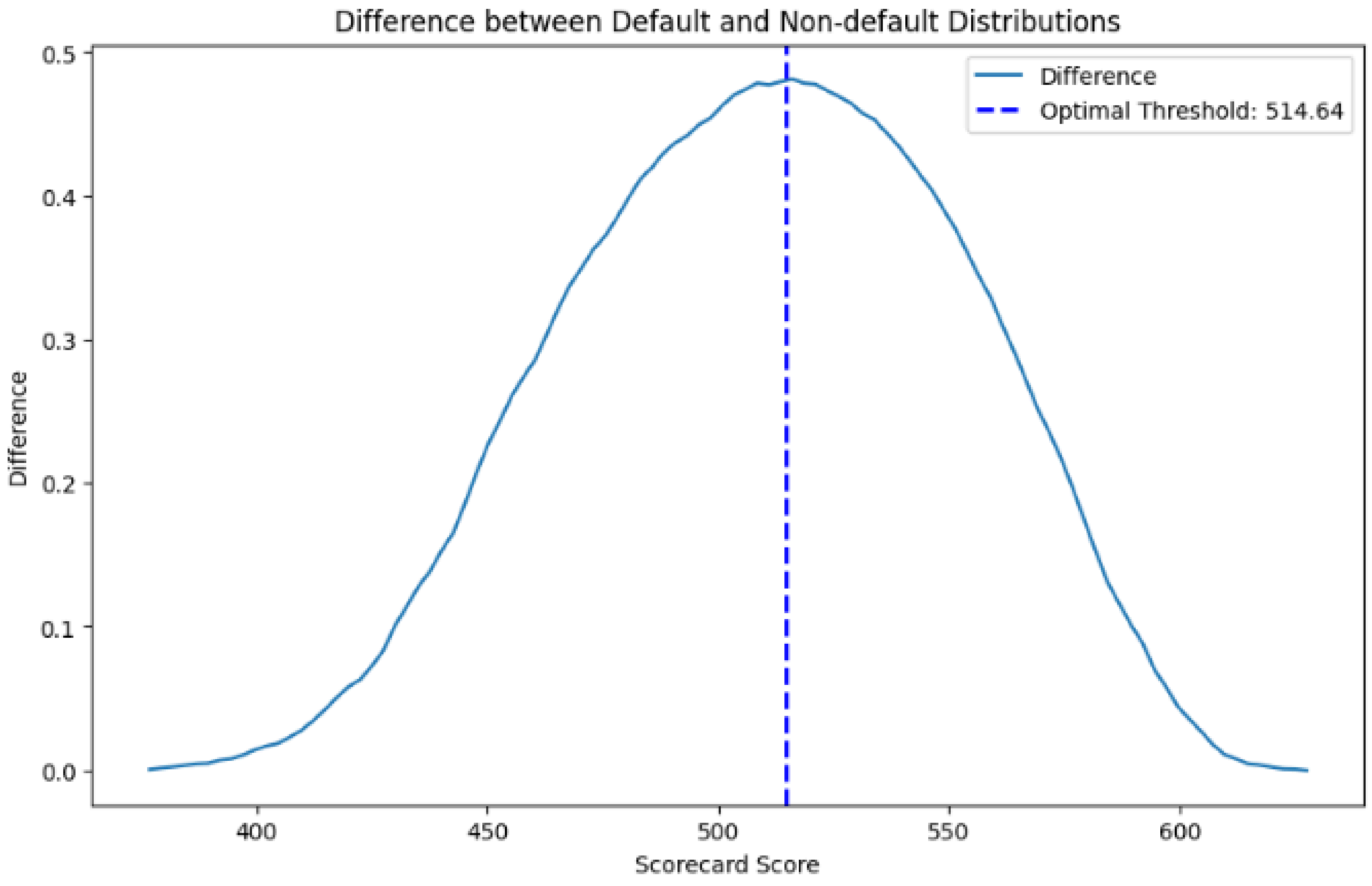
The peak of the non-default distribution is at higher scores, around 550, indicating that most non-default events are concentrated in the higher score ranges. Conversely, the peak of the default distribution is at lower scores, around 450, suggesting that most default events are concentrated in the lower score ranges.

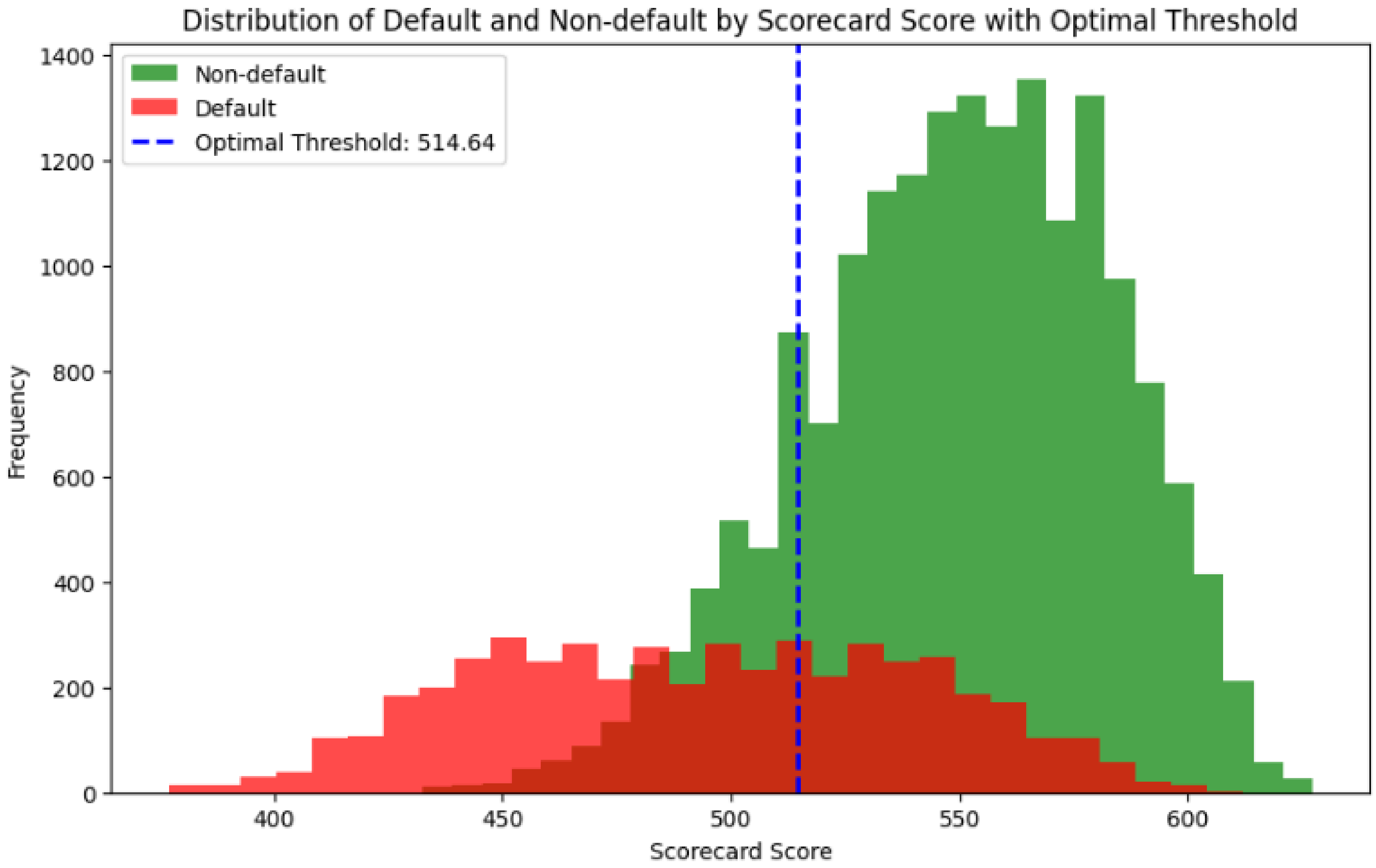
cut off score analyze

first method : maximizes the difference between the two distributions

- Separate scores into default and non-default groups.
- Approximate the CDFs of both groups using histograms and cumulative sums.
- Calculate the absolute difference between the two CDFs at each score bin.
- Choose the score value that maximizes this difference as the optimal threshold for classification.

result


Difference between Default and Non-default Distributions
Difference
Optimal Threshold: 514.64
0.0
0.1
0.2
0.3
0.4
0.5
400
450
500
550
600
Scorecard Score


Distribution of Default and Non-default by Scorecard Score with Optimal Threshold
Non-default
Default
Optimal Threshold: 514.64
Frequency
0
200
400
600
800
1000
1200
1400
400
450
500
550
600
Scorecard Score

	Score Threshold	Cumulative Bad Rate	Cumulative Good Rate	Acceptance Rate
0	300	21.69%	78.31%	100.00%
1	400	21.44%	78.56%	99.75%
2	500	10.01%	89.99%	81.79%
3	514	7.98%	92.02%	74.34%
4	550	2.85%	97.15%	43.43%
5	600	0.05%	99.95%	3.63%

and when we look at bad rate

↓

- **The Cumulative Bad Rate** is calculated as the number of defaulted loans with a score equal to or above the given threshold, divided by the total number of loans (both defaulted and non-defaulted) in the population. **It represents the expected default rate if you were to approve all loan applications with scores at or above that threshold.**
- So, suppose you were to set a threshold of, say, 514 (optimal threshold before). In that case, This score can be considered the optimal cutoff score because it strikes the best balance between accepting a reasonable number of loans (73.74%) and keeping the bad rate relatively low (7.86%).

	loan_amnt	loan_int_rate	profit_loss	loan_status
0	35000	0.16	-35000.00	1
1	1000	0.11	111.40	0
2	5500	0.13	-5500.00	1
3	35000	0.15	-35000.00	1
4	35000	0.14	-35000.00	1
...	...	...	...	...
32576	5800	0.13	763.28	0
32577	17625	0.07	1320.11	0
32578	35000	0.11	-35000.00	1
32579	15000	0.11	1722.00	0
32580	6475	0.10	646.85	0

second method : profit

To calculate the profit or loss for each loan, we consider two scenarios:

- profit_loss=loan_amnt*(1+loan_int_rate)-loan_amnt **(if loan does not default)**
- profit_loss=-loan_amnt **(if loan defaults)**

profit in each threshold

	Score Threshold	Cumulative Bad Rate	Acceptance Rate	Profit
0	350.00	21.82%	100.00%	-54463425.11
1	400.00	21.57%	99.76%	-53688950.12
2	450.00	17.19%	95.22%	-35332310.50
3	500.00	10.09%	81.83%	-9381296.51
4	514.64	7.86%	73.74%	-5082928.43
5	537.00	4.62%	56.13%	-125748.83
6	538.00	4.48%	55.31%	14842.43
7	550.00	2.85%	43.50%	2292758.42
8	600.00	0.04%	3.59%	767715.67
9	650.00	0.00%	0.00%	0.00

At a score threshold of 538.0, the data shows a least positive profit of $14,842.43 and the acceptance rate is 55.31%. This threshold represents the highest score value, relatively low acceptance

strategy

Score Threshold of 538+ (customers with below this score will got reject).

choosing a score threshold of 538.0 would be an appropriate strategy in situations where the lender prioritizes a low-risk portfolio and maintaining a positive profit, even if modest. This approach may be suitable in the following scenarios:

Economic Downturn or Recessionary Environment lenders may adopt a more conservative stance to mitigate potential losses

Risk-Averse Lending Approach: Some lenders may have a lower tolerance for risk and prefer to maintain a high-quality loan portfolio

Score Threshold of 514+

On the other hand, choosing a score threshold of 514 would be a suitable strategy when the lender aims to capture a larger market share and is willing to accept a higher level of risk This approach may be advantageous in the following situations:

Aggressive Market Expansion: If the lender operates in a highly competitive market and seeks to gain a significant market share, a lower score cutoff like 514.64 can help approve more loan applications.

High Risk Tolerance: Lenders with a higher tolerance for risk may be willing to accept the potential losses associated with a lower score threshold like 514.64